

ほぼ周期文字列

writer : nuip

tester : zerokugi

1

概要

- 文字列が与えられる
- クエリで与えられる部分文字列と周期 t について、一文字変えて周期 t にできるか判定せよ

問題の言い換え

- S が周期 t を持つ $\Leftrightarrow S[1 \dots |S| - t] = S[t \dots |S|]$

証明の方針

S が周期 t を持つ $\Leftrightarrow \forall i = 1, \dots, |S| - t \ S[i] = S[i + t]$ であるため、

$\forall i = 1, \dots, |S| - t \ S[i] = S[i + t] \Leftrightarrow S[1 \dots |S| - t] = S[t \dots |S|]$ を示す。

\Rightarrow は対応するそれぞれの文字が同じなので、明らか。

\Leftarrow も、もともとと同じ所からとってきた文字は同じであることを使えば示せる。雰囲気としては、次の通り

$\forall i = 1, \dots, |S| - t \quad S[i] = S[i + t] \iff S[1 \dots |S| - t] = S[t \dots |S|]$ の理由

長さ t の文字列を大文字のアルファベットで表す。
 $S[0 \dots |S| - t]$ が ABCD というかたちをしていて、
 $S[t \dots |S|]$ が EFGH というかたちをしているとする。

ABCD_
S
_EFGH

$\forall i = 1, \dots, |S| - t \quad S[i] = S[i + t] \quad \Leftarrow \quad S[1 \dots |S| - t] = S[t \dots |S|]$ の理由

- 同じ位置にあるので、A=B

ABCD_

S

_ABCD

$\forall i = 1, \dots |S| - t \ S[i] = S[i + t] \iff S[1 \dots |S| - t] = S[t \dots |S|]$ の理由

- 同じ位置にあるので、A=C

AACD_

s

_AACD

$\forall i = 1, \dots |S| - t \quad S[i] = S[i + t] \quad \Leftarrow \quad S[1 \dots |S| - t] = S[t \dots |S|]$ の理由

- 同じ位置にあるので、A=D

AAAD_

S

_AAAD

$\forall i = 1, \dots, |S| - t \quad S[i] = S[i + t] \quad \Leftarrow \quad S[1 \dots |S| - t] = S[t \dots |S|]$ の理由

- よって、 S はAAAAAという形をしている。つまり周期 t である。

AAAA_

S

_AAAA

問題の言い換え

- S が周期 t を持つ $\Leftrightarrow S[1 \dots |S| - t] = S[t \dots |S|]$

つまりこの問題は、1文字まで変えていい時与えられた部分文字列どうしが等しくなるようにできるか判定する問題。

解法

- 部分文字列どうしの比較は高速にできる
 - ハッシュを計算する
 - ローリングハッシュ
 - Suffix array

解法

- 一文字変えて $S[l, \dots, r] = S[L \dots R]$ にできるか?
- $l < L$ とする
- 最初から同じなら Yes
- 交わってる部分 $[l, r] \cap [L, R] = [L, r]$ を変える場合と、そうでない場合で場合分け
- (この部分については、頑張って解説を読むよりも、自分で考察したほうが分かりやすいと思います。)

交わってない部分を変える

- 一文字違いまでなら、そこを変えれば等しくなる。
- そうでなければ2文字以上変えなきゃだめ。

交わってる部分を変える

できる例

2箇所違う部分がある。

a b c a x c a b c

a b c a x c a b c

真ん中のxをbにすればよい

交わってる部分を変える

できない例

2箇所違う部分がある。

abcaxcayc

abcaxcayc

真ん中のxをbに変えてもyに変えてもどちらかは一致しない。

交わってる部分を変える

i 文字目を変更する場合、 $S[i - t] = S[i + t]$ でないといけない。

このとき、その2箇所以外に異なっている文字が存在しなければ、 i 文字目を $S[i - t]$ で書き換えれば良い。

それ以外ならできない。

別解法

- ややこしい条件を考えなくても...
- ローリングハッシュ($S[l] \times r^{r-l} + \dots + S[r] \times r^0 \pmod M$)を使っている時
 - (右から) x 文字目を c から c' に変えた場合のハッシュが求められる
 - $(c' - c) \times r^x$ を足すだけ
- 「ここを変えないと絶対等しくすることはできない」という文字 (例:一番右にある異なる文字) を変えてみて、全体が等しくなるか試せば良い