

Skirtumas

Pažymėkime duotąjį žodį W (jo ilgis n).

Kai X ir Y yra mažosios lotyniškos raidės (žymėsime tai $X, Y \in \mathcal{A}$, $\mathcal{A} = \{a, \dots, z\}$), tegul

$$f(X, Y) = \max \{I(X) - I(Y) : I \text{ yra } W \text{ fragmentas ir } I(Y) > 0\},$$

kur $I(X)$ žymi, kiek kartų raidė X pasikartoja fragmente I . Tuomet uždavinio atsakymas yra

$$R = \max \{f(X, Y) : X, Y \in \mathcal{A}\}.$$

Kiekvienai žodžio W raidei apskaičiuokime, kur tame žodyje yra artimiausia tokia pati raidė iš kairės ir iš dešinės. Visą šią informaciją galime apskaičiuoti vienu ypu atlikdami $O(n)$ operacijų. Kai $i = 1, 2, \dots, n$, tegul $k[i]$ ir $d[i]$ žymi, kur yra artimiausia identiška raidė i -ajai W raidei, atitinkamai iš kairės ir iš dešinės.

Užfiksukime X ir Y . Su kiekvienu $j = 0, 1, \dots, n$ tegul $T[j]$ žymi, kiek kartų raidė X kartojasi tarp pirmųjų j žodžio W raidžių. Dabar galime apskaičiuoti, kiek kartų X kartojasi bet kuriame W fragmente: jei I yra fragmentas nuo a -osios iki b -osios W raidės imtina (rašykime $I = [a, b]$), tai jame bus $T[b] - T[a - 1]$ raidžių X . Visus $T[j]$ galime apskaičiuoti atlikdami $O(n)$ operacijų.

Apskaičiuodami $f(X, Y)$ mes nagrinėjame tik tuos fragmentus, kuriuose yra bent viena raidė Y , todėl paeiliui renkamės visas žodyje W esančias raides Y kaip fragmento „pradžią“. Mums apsimoka žiūrėti tik į tokius fragmentus $[a, b]$, kad $a = 1$ arba $W[a - 1] = Y$ ir $b = n$ arba $W[b + 1] = Y$.

Pavyzdžiui, jei W būtų tokios formos ir kaip fragmento „pradžią“ pasirinktume rodykle pažymėtą Y

$$YY***\overset{\downarrow}{Y}*Y**,$$

tai visi šeši mus dominantys fragmentai būtų tokie:

$$\begin{array}{ll} \underline{YY***Y}*Y** & \underline{YY***Y}*Y** \\ Y\underline{Y***Y}*Y** & Y\underline{Y***Y}*Y** \\ \underline{YY***Y}*Y** & \underline{YY***Y}*Y** \end{array}$$

Naudojame dinaminį programavimą. Kiekvienam j apskaičiuosime dvi vertes $U[j]$ ir $V[j]$, kurios reiškia:

- $U[j]$ žymi didžiausią galimą $I(X) - I(W[j])$ vertę, kai $I = [i, j]$, $i \leq j$;
- $V[j]$ žymi didžiausią dalimą $I(X) - I(W[j])$ vertę, kai $I = [j, i]$, $i \geq j$.

Masyvus U ir V galime užpildyti atlikdami $O(n)$ operacijų, jei naudosime rekursines formules

$$\begin{aligned}U[j] &= T[j] - T[k[j]] - 1 + \max\{0, U[k[j]]\} \\V[j] &= T[d[j] - 1] - T[j - 1] - 1 + \max\{0, V[d[j]]\}.\end{aligned}$$

Galiausiai, jei pasirinkome p kaip „pradžią“ fragmentui, tai didžiausias skirtumas tokiam fragmentui yra

$$U[p] + V[p] \quad (+1, \text{ jei } X \neq W[p]).$$

Iš tikrųjų mums Y užfiksuoti nereikia: užfiksuojame tik X ir kai pasirenkame „pradžią“ p (ją renkamės paeiliui nuo 1 iki n), tai tam žingsniui įsivaizduojame, kad $Y = W[p]$. Taip su kiekvienu užfiksuotu X apskaičiuojame $\max\{f(X, Y) : Y \in \mathcal{A}\}$ atlikdami $O(n)$ operacijų. Taigi visas algoritmo sudėtingumas yra $O(|\mathcal{A}|n) = O(n)$.