Merkkijonomenetelmät (syksy 2003)

Lisätehtäviä

Tehtävistä saa lisäpisteitä (1 piste per tehtävä, mutta *yhteensä korkeintaan 4*), jos palauttaa vastaukset *kirjallisena* luennoitsijalle tiistain (2.12) luennon alussa. Lisäpisteet lasketaan puuttuvien laskuharjoituspisteiden päälle (jos ennen oli X tehtävää tehtynä Y:stä, lisätehtävien jälkeen voi olla $\min(X+4, Y)/Y$). Luennolla käydään läpi tehtävien ratkaisuja, ja muutenkin kerrataan kurssin sisältöä.

1. Anna lineaariaikainen algoritmi Boyer-Moore:n δ_2 taulukon muodostamiseksi. (*Vihje.* Löytyy kurssikirjasta. TAI. Käytä eliminointimenetelmän Witness taulukkoa W hyväksi: Muodosta käänteisjonon Witness taulukko. Tämä kääntyy sopivalla muunnoksella δ_2 taulukoksi. Tapaus jossa siirto vie hahmon ohi epätäsmäyskohdan pitää käsitellä erikseen. Tämä erikoistapaus hoituu Knuth-Morris-Pratt:in $fail$-taulukon avulla: Tarkastele $fail[m+1]$:stä alkavaa linkkilistaa.)

2. Merkkijonon q-piirre on sen q:n mittainen osajono. Tehtävänä on muodostaa taulukko $Occ[1 \dots \sigma^q]$, johon on talletettu jokaiselle q-piirteelle linkkilista sen esiintymäkohdista tekstissä S. Näytä, että tämä voidaan tehdä ajassa $O(\sigma^q + n)$, missä σ on aakkoston koko ja n on tekstin S pituus (triviaalialgoritmi veisi ajan $O(\sigma^q + qn)$). (*Vihje.* Käytä samaa tekniikkaa kuin Rabin-Karp-algoritmissa.)

3. Kurssilla ei käsitelty lainkaan aihetta "Likimääräinen hahmonsovitus staattisissa jonoissa". Keksitkö miten kurssilla esitettyjä asioita yhdistelemällä saadaan johdettua algoritmeja tähän ongelmaan. Anna ainakin kaksi erilaista lähestymistapaa. Yksityiskohtia ei tarvitse selittää.

4. Tutustu Kappaleeseen 3.5.1. Ratkaisussa B lasketaan taulukko $maxPrefix[1 \dots n]$ käyttäen hahmon loppuosatrietä. Tämän tilalla voitaisiin käyttää hahmon loppuosapuuta. Miten algoritmi muuttuu? Mikä on sen aikavaatimus? (*Vihje:* loppuosapuussa ei ole loppuosalinkkejä implisiittisille solmuille, joten pitää käyttää $Canonize$-funktiota aivan kuten loppuosapuun muodostusalgoritmissa. Aikavaativuutta voi analysoida kuten muodostuksen aikavaativuutta.)

5. Annettuna pelkkä loppuosataulukko $Pos[1 \dots n]$. Miten muodostetaan jokin merkkijono S, jonka loppuosataulukko Pos on? Jonon S tulisi olla mahdollisimman pienestä aakkostosta (ongelma on triviaali jos valitsee aakkoston kooksi luvun n). (*Vihje:* Käsittele loppuosat järjestyksessä $Pos[1], Pos[2], \dots$. Valitse kunkin loppuosan $Pos[i]$ ensimmäiseksi merkiksi uusi merkki vain jos taulukko tulisi muuten ristiriitaiseksi tutkimalla loppuosien $Pos[i-1]+1$ ja $Pos[i]+1$ järjestystä taulukossa (näiden sijainnit löytyvät taulukon $Reverse$ avulla).)

6. Taulukossa $A[1 \dots n]$ on talletettu merkkejä aakkosjärjestyksessä siten, että sama merkki voi toistua. Taulukko vie $O(n \log \sigma)$ bittiä tilaa, missä σ on aakkoston koko. Näytä, että taulukko voidaan tiivistää tilaan $n + o(n) + O(\sigma \log \sigma)$ bittiä siten, että positiossa i oleva merkki voidaan selvittää vakio-ajassa. (*Vihje:* Käytä bittivektorin $rank$-operaatiota.)

7. *VASTAA KURSSIKYSELYYN* (http://www.cs.helsinki.fi/kurssit/kyselyt/).