Merkkijonomenetelmät (syksy 2003)

Harjoitus 9

Ryhmä	aika	paikka
Veli Mäkinen	to 20.11 klo 8.30–10	A319
Anna Pienimäki	pe 21.11 klo 8.30–10	B453

1. Anna algoritmi, joka löytää ajassa $O(m+n)$ merkkijonojen $A = a_1a_2\cdots a_m$ ja $B = b_1b_2\cdots b_n$ pisimmän yhteisen osajonon, eli pisimmän jonon, joka esiintyy sekä A:ssa että B:ssä. (*Vihje:* Muodosta loppuosapuu merkkijonolle $a_1a_2\cdots a_m\#b_1b_2\cdots b_n\$$, missä # on lisämerkki, joka ei esiinny A:ssa eikä B:ssä.)

2. Muodosta loppuosataulukko jonolle `LALLILLA`, ja sovella siihen binäärihakumetodia löytääksesi hahmon `ALLI`.

3. Karp-Miller-Rosenberg menetelmässä, siirryttäessä nimennästä $Name_{p/2}$ nimentään $Name_p$, pitää järjestää joukko nimipareja siten, että parit ovat nousevassa järjestyksessä ensisijaisesti ensimmäisen arvon mukaan ja toissijaisesti toisen arvon mukaan. Esimerkiksi siirryttäessä Kuvan 3.4 (sivu 102) $Name_1$ nimennästä nimentään $Name_2$ tulee järjestettäväksi jono $(1,2)(2,1)(1,1)(1,2)(2,2)(2,1)(1,1)(1,-1)$. Järjestettynä tämä on $(1,-1)(1,1)(1,1)(1,2)(1,2)(2,1)(2,1)(2,2)$, ja uusi nimentä on $1 = (1,-1), 2 = (1,1), 3 = (1,2), 4 = (2,1), 5 = (2,2)$. Koska nimiparien $\{(i,j)\}$ kummatkin komponentit ovat kokonaislukuja, $1 \leq i \leq n$, $-n \leq j \leq n$, voidaan parit järjestää lineaarisessa ajassa käyttäen taulukkojärjestämistä. Miten?

4. Jatkoa edelliskerran tehtävälle 2. Tehtävän tarkoituksena on näyttää, että $O(m \log m)$ esiprosessointiaika riittää vastaamaan kysymykseen $prefix(i', i)$, eli antamaan hahmon loppuosien $P_{i'\ldots m}$ ja $P_{i\ldots m}$ pisimmän yhteisen alkuosan pituuden. Huomataan seuraava yhteys *hahmolle rakennettuun* loppuosataulukkoon (käyttäen Kappaleen 3.4.4 merkintöjä)

$$prefix(i', i) = \min_{j' < k \leq j} LCPprev[k], \tag{1}$$

missä $j' = Reverse[i']$ ja $j = Reverse[i]$, ja oletetaan että $j' < j$ (jos näin ei ole, vaihdetaan rajojen järjestystä). Näytä miten $LCPprev$-taulukko (tai itse asiassa mikä tahansa taulukko) voidaan esiprosessoida $O(m \log m)$ ajassa siten, että minkä tahansa indeksivälin $[j', j]$ minimi voidaan tulostaa vakioajassa. (*Vihje:* Talleta kullekin i indeksivälien $\{i \ldots i + 2^k - 1 \mid 0 \leq k \leq \lceil \log m \rceil\}$ minimit. Sopivassa järjestyksessä taulukoimalla kukin arvo voidaan laskea vakioajassa. Lopuksi huomaa, että minkä tahansa kyselyvälin minimi voidaan lukea taulukosta ottamalla minimi kahdesta taulukkoon talletusta arvosta.)