Merkkijonomenetelmät (syksy 2003)

Harjoitus 6

Ryhmä	aika	paikka
Veli Mäkinen	to 30.10 klo 8.30–10	A319
Anna Pienimäki	pe 31.10 klo 8.30–10	B453

1. Näytä miten täsmäävien merkkiparien joukko $M = M(A, B) = \{(i, j) \mid a_i = b_j\}$ voidaan muodostaa ajassa $O(\sigma+m+n+|M|)$, missä $A = a_1 \cdots a_m$ ja $B = b_1 \cdots b_n$ ovat aakkoston $\Sigma = \{1, 2, \ldots \sigma\}$ merkkijonoja.

2. Simuloi harvan dynaamisen ohjelmoinnin (Algoritmi 2.4.8) toimintaa etäisyyden $D'(A, B)$ laskemiseksi syötteellä $A =$ `tukholma` ja $B =$ `stockholm`. Tietorakenteen $\mathcal{T}$ toimintaa ei tarvitse simuloida.

3. Johda harvaan dynaamiseen ohjelmointiin perustuva algoritmi edellisten harjoitusten tehtävälle 5, eli anna algoritmi, joka tulostaa lyhyimmän sellaisen tekstin T:n osajonon pituuden, joka sisältää hahmon P alijononaan. Algoritmin tulisi toimia ajassa $O(|M|)$, kun täsmäävien merkkiparien joukko $M = M(P, T)$ on annettu.

4. Yleisen editointietäisyyden $D_G(A, B)$ laskemisen yhteydessä käydään läpi etäisyysmatriisin (d_{ij}) jokaisen alkion kohdalla koko operaatiojoukko G. Esitä tehokkaampi tapa käydä läpi alkion $d_{i,j}$ kohdalla kysymykseen tulevat G:n alkiot. Mikä on parannetun menetelmän aikavaatimus? (*Vihje:* Rakenna kaksi Aho-Corasick-automaattia sanoille, jotka esiintyvät G:n jonkin operaation oikeana/vasempana puolena.)

5. Lyhimmän kiertotien menettelyssä tuplausalgoritmi aloitetaan arvolla $x = \lfloor \frac{n-m}{2} \rfloor + 1$. Algoritmi toimisi oikein vaikka tuplaus aloitettaisiin hieman luonnollisemmalla aloitusarvolla kuten $x = 1$. Päästäänkö tällä aloitusarvolla samaan aikavaatimukseen kuin alkuperäisellä? Perustele.

HUOM. Näissä harjoituksissa jaetaan harjoitustyöaiheet. Lisätietoa löytyy sivulta `http://www.cs.helsinki.fi/u/vmakinen/mjm-s03/projekti.html`