Merkkijonomenetelmät (kevät 2007)

Harjoitus 3

Jarkko Toivonen pe 9.2. klo 10.15–12 CK107

1. Kuvaa sellainen merkkijonoparien perhe (A_i, B_i), $i = 1, 2, \ldots$, että $|A_i| = |B_i| < |A_{i+1}|$ ja editointietäisyyttä $D(A_i, B_i)$ vastaa "paljon" erilaisia operaatiojonoja. Löydätkö esimerkin, jossa erilaisten operaatiojonojen määrä kasvaa oleellisesti nopeammin kuin $|A_i|$?

2. Oletetaan, että editointioperaatioista on käytössä poisto ja lisäys mutta ei muutosta. Tällöin dynaamisen ohjelmoinnin pohjana oleva palautuskaava saa muodon

$$
\begin{aligned}
d_{0j} &= j, \qquad 0 \le j \le n, \\
d_{i0} &= i, \qquad 1 \le i \le m, \\
d_{ij} &= \min \begin{cases} d_{i-1,j-1} + (\text{if } a_i = b_j \text{ then } 0 \text{ else } \infty) \\ d_{i-1,j} + 1 \\ d_{i,j-1} + 1 \end{cases}, \quad 1 \le i, j \le n.
\end{aligned}
$$

   Osoita, että lävistäjälemma (Lemma 2.3.5) pätee nyt muodossa

$$d_{ij} = d_{i-1,j-1} \text{ tai } d_{ij} = d_{i-1,j-1} + 2$$

3. Jono B on jonon A *alijono*, jos B saadaan poistamalla A:sta nolla tai useampia merkkejä. Olkoon P ja T kaksi merkkijonoa (hahmo ja teksti). Anna algoritmi, joka tulostaa lyhyimmän sellaisen tekstin T osajonon pituuden, joka sisältää hahmon P alijononaan. Esim. syötteellä $P =$`aba` ja $T =$`aacbacbca` algoritmin tulisi tulostaa 4 (osajonon <u>a</u>c<u>ba</u> pituus). Algoritmin tulisi toimia ajassa $O(mn)$, missä $m = |P|$ ja $n = |T|$.

4. (a) Löydä rajoitetulla dynaamisella ohjelmoinnilla (Algoritmi 2.3.6) tekstistä $S =$ `CAGATAAGAGAA` hahmon $P =$ `GATAA` likimääräiset esiintymät, joissa on korkeintaan yksi virhe (eli $k = 1$).

   (b) Simuloi Myersin algoritmin (Algoritmi 2.4.7) toimintaa, kun se laskee sarakkeen $j = 6$ (a)-kohdan syötteellä.

5. Valitse aihe kurssin ohjelmointi-projektiin. Tästä tehtävästä saavat merkinnän kaikki, jotka ovat valinneet aiheensa aikaisemmin tai valitsevat sen viimeistään näissä harjoituksissa.