

Merkkijonomenetelmät (kevät 2008)

Harjoitus 3

Niko Välimäki ti 23.09. klo 10.15–12 CK111

1. Laske jonojen BANAANI ja ANANAS editointietäisyys.
2. Kuvaa sellainen merkkijonoparien perhe (A_i, B_i) , $i = 1, 2, \dots$, 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 $|A_i|$?
3. Oletetaan, että editointioperaatioista on käytössä poisto ja lisäys mutta ei muutosta. Tällöin dynaamisen ohjelmoinnin pohjana oleva palautuskaava saa muodon

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

Osoita, että lävistäjälemma (Lemma 2.3.5) pätee nyt muodossa

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

4. 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 $P = \text{aba}$ ja $T = \text{aacbacbca}$ algoritmin tulisi tulostaa 4 (osajonon acba pituus). Algoritmin tulisi toimia ajassa $O(mn)$, missä $m = |P|$ ja $n = |T|$.
5. (a) Löydä rajoitetulla dynaamisella ohjelmoinnilla (Algoritmi 2.3.6) tekstistä $S = \text{CAGATAAGAGAA}$ hahmon $P = \text{GATAA}$ 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.