Merkkijonomenetelmät (kevät 2005)

Harjoitus 4

Pasi Rastas ma 28.2 klo 10.15–12 C221

1. Todista, että editointietäisyys $D(A, B)$ on metriikka (ks. luentomuistiinpanot s. 32).

2. (a) Muodosta matriisi (d_{ij}) jonoille `tokio` ja `kioto`.
   (b) Muodosta matriisi (f_{kp}) samoille jonoille (esim. matriisista (d_{ij}) tai simuloimalla luentojen algoritmia).
   (c) Esitä matriisissa (d_{ij}) kaikki optimipolut d_{mn}:stä d_{00}:aan. Montako niitä on?
   (d) Valitse ylintä reittiä kulkeva optimipolku ja esitä vastaava jonojen kohdistus sekä vastaava polku (f_{kp}):ssä.

3. 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 $|A_i|$?

4. Yleisen editointietäisyyden $D_G(A, B)$ laskevassa algoritmissa 2.4.10 eniten aikaa kuluu rivillä 6, missä tutkitaan, esiintyykö x $A_{1\ldots i}$:n ja y $B_{1\ldots j}$:n lopussa. Algoritmia voidaan nopeuttaa etsimällä etukäteen x:n esiintymät A:ssa ja y:n esiintymät B:ssä kaikille $x \to y \in G$. Täydennä tarvittavat yksityiskohdat ja analysoi parannetun algoritmin aikavaatimus.

5. Osoita, että (poisto/lisäys)-editointietäisyydelle lävistäjälemma pätee muodossa

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