Merkkijonomenetelmät (kevät 2005)

Harjoitus 5

Pasi Rastas ma 7.3 klo 10.15–12 C221

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

2. *Episodihahmonsovituksessa* tehtävänä on löytää tekstin S minimaaliset osajonot, jotka sisältävät hahmon P alijononaan. Osajono on minimaalinen, jos yhdenkin merkin pudottaminen alusta tai lopusta merkitsisi, että jono ei enää sisällä hahmoa alijononaan. Esim. syötteellä $S =$ `aacbacbca` ja $P =$ `aba` tulisi löytää osajonot $S_{2\cdots5} =$ `acba` ja $S_{5\ldots9} =$ `acbca`. Anna harvaan dynaamiseen ohjelmointiin perustuva algoritmi episodihahmonsovitukseen. Algoritmin tulisi toimia ajassa $\mathcal{O}(|M|)$, kun täsmäävien merkkiparien joukko M on annettu käänteisessä sarakejärjestyksessä. (Vihje: Pidä kirjaa hahmon jokaisen alkuosan viimeisimmän episodiesiintymän alkukohdasta.)

3. (a) Löydä tehostetulla dynaamisella ohjelmoinnilla (Algoritmi 2.7.6) tekstistä $S =$ `CAGATAAGAGAA` hahmon $P =$ `GATAA` likimääräiset esiintymät, joissa on korkeintaan yksi virhe (eli $k = 1$).

   (b) Simuloi Myersin algoritmin (Algoritmi 2.7.14) toimintaa, kun se laskee sarakkeen $j = 6$ (a)-kohdan syötteellä.

4. Olkoon (d'_{ij}) (poisto/lisäys)-editointietäisyyden laskennassa syntyvä matriisi (luentojen luku 2.5). Merkitään $\Delta v'_{ij} = d'_{ij} - d'_{i-1,j}$ ja $\Delta h'_{ij} = d'_{ij} - d'_{i,j-1}$.

   (a) Osoita, että kaikilla i ja j pätee $\Delta v'_{ij}, \Delta h'_{ij} \in \{-1, 1\}$. (Voit käyttää harjoitusten 4 tehtävän 5 tulosta.)

   (b) Myös (poisto/lisäys)-editointietäisyyden tapauksessa päästään etäisyyden laskennasta likimääräiseen hahmonsovitukseen muuttamalla rivin 0 alustusta kuten luentojen luvussa 2.7. Pätevätkö tällöin kohdan (a) ja harjoitusten 4 tehtävän 5 tulokset? Perustele.

5. Esitä luentojen lemmaa 2.7.10 vastaava sääntö tehtävän 4(a) $\Delta v'$- ja $\Delta h'$-arvojen laskemiseen ja todista se.