Merkkijonomenetelmät (kevät 2005)

Harjoitus 6

Pasi Rastas ma 14.3 klo 10.15–12 C221

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

2. Merkkijonon q-*gramma* on sen q:n pituinen osajono. Esim. jonon `ararat` 3-grammat ovat `ara`, `rar`, `ara` ja `rat`.

   (a) Osoita, että jos A ja B ovat merkkijonoja, joiden editointietäisyys $D(A, B)$ on enintään k, niin niillä on ainakin $|A| - q + 1 - kq$ yhteistä q-grammaa.

   (b) Hahmottele (a)-kohdan tulokseen perustuva seulontamenetelmä likimääräiseen hahmonsovitukseen.

3. Havainnollista Galil-Giancarlo-algoritmin toimintaa, kun teksti on `abaabaaaa`, hahmo on `abaaaa`, ja epätäsmäysten määrän yläraja on $k = 1$.

4. Olkoon R monijoukko, jonka koko on n ja jossa on $k < n$ eri arvoa. Osoita, että ternääripikajärjestäminen järjestää joukon R ajassa $\mathcal{O}(n \log k)$ olettaen, että jakoalkioksi valitaan aina mediaani. (Vihje: Käytä tasoitettua analyysiä, missä kunkin kolmeenjaon maksavat joukkojen $R_=$ ja R' alkiot, missä R' joukoista $R_<$ ja $R_>$ se, joka sisältää vähemmän eri arvoja.)

5. Esitä luentojen Algoritmista 3.3.2 (lopusta alkuun kantalukujärjestäminen) yleistetty versio, joka toimii myös eripituisilla merkkijonoilla lauseen 3.3.4 mukaisessa ajassa $\mathcal{O}(||\mathcal{R}|| + m\sigma)$, missä $||\mathcal{R}||$ on jonojen kokonaispituus ja m on pisimmän jonon pituus.