

Merkkijonomenetelmät (syksy 2008)

Harjoitus 4

Jarkko Toivonen ti 30.2. klo 10.15–12 CK111

1. Merkkijonon *q-gramma* on sen q :n pituinen osajono. Esim. jonon **ararat** 3-grammat ovat **ara**, **rar**, **ara** ja **rat**. 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.
2. Hahmottele tehtävän 1 tulokseen perustuva seulontamenetelmä likimääräiseen hahmonsovitukseen.
3. Luennoilla osoitettiin, että yleisellä vertailuihin perustuvalla algoritmilla merkkijonojen järjestäminen vaatii ajan $\Omega(n \log n \log_{\sigma} n)$ parhaassakin tapauksessa. Mikä on *pahimman* tapauksen aikavaatimus?
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' on joukoista $R_{<}$ ja $R_{>}$ se, joka sisältää vähemmän eri arvoja.)
5. Valitse aihe kurssin ohjelmointi-projektiin. Tästä tehtävästä saavat merkinnän kaikki, jotka ovat valinneet aiheensa aikaisemmin tai valitsevat sen viimeistään näissä harjoituksissa.