## Merkkijonomenetelmät (kevät 2007)

### Harjoittelutehtäviä

1. VASTAA KURSSIKYSELYYN (`http://ilmo.cs.helsinki.fi/kurssit/servlet/Valinta`).

2. Mitä yhteyksiä/eroavaisuuksia on seuraavilla käsitepareilla?

   (a) Knuth-Morris-Pratt-algoritmi ja Aho-Corasick-algoritmi.

   (b) Shift-Or-algoritmi ja Myersin bittirinnakkainen algoritmi.

   (c) Karp-Rabin-algoritmi ja Karp-Miller-Rosenberg nimentätekniikka.

   (d) Editointietäisyyden laskeminen ja likimääräinen hahmonsovitus.

   (e) Merkkijonopikajärjestäminen ja ternääripuu.

   (f) Lopusta-alkuun ja alusta-loppuun kantalukujärjestäminen.

   Muutama lause riittää vastaukseksi kuhunkin kohtaan

3. Anna Aho-Corasick-automaatti hahmojoukolle $\{\mathtt{ARI}, \mathtt{ARMAS}, \mathtt{KARITA}, \mathtt{MARKO}, \mathtt{RITVA}\}$. Simuloi automaatin käyttöä tekstillä $T = \mathtt{SANKARITAR}$.

4. Olkoon $S$ $n$:n pituinen merkkijono ja $R$ $m$:n merkin monijoukko vakioaakkostossa $\Sigma$. $R$:n esiintymä $S$:ssä on $m$:n pituinen osajono, joka koostuu täsmälleen $R$:n merkeistä. Esimerkiksi joukon $R = \{\mathtt{A}, \mathtt{A}, \mathtt{B}, \mathtt{C}\}$ ainoa esiintymä jonossa $S = \mathtt{ABAHGCABAH}$ on $\mathtt{CABA}$. Anna ajassa $\mathcal{O}(n+m)$ toimiva algoritmi, joka löytää kaikki $R$:n esiintymät $S$:ssä.

5. Merkkijonoja $A = a_1a_2\cdots a_m$ ja $B = b_1b_2\cdots b_n$ kutsutaan *numeerisiksi merkkijonoiksi*, jos aakkostona on reaaliluvut, eli $a_i, b_j \in \mathbb{R}$ kaikilla $1 \leq i \leq m$, $1 \leq j \leq n$. Tarkastellaan erästä numeeristen merkkijonojen editointietäisyyttä, $D_K(A, B)$, joka määritellään seuraavasti. Sallitaan kolme editointioperaatiota: korvaus, poisto ja lisäys. Korvausoperaation $a_i \to b_j$ hinta on $|a_i - b_j|$. Poistot $a_i \to \epsilon$ ja lisäykset $\epsilon \to b_j$ ovat *ilmaisia*, mutta niiden yhteismäärä saa olla korkeintaan $K$. Editointietäisyys $D_K(A, B)$ on siis pienin kokonaishinta editointijonolle, joka muuttaa $A$:n $B$:ksi ja sisältää enintään $K$ lisäystä ja poistoa. Tehtävänä on kehittää etäisyyden $D_K(A, B)$ laskemiseksi algoritmi, joka toimii ajassa $\mathcal{O}(mnK)$.

6. Kuvaa periaatetasolla menetelmiä, joilla binäärihakua voidaan tehostaa, kun alkiot ovat merkkijonoja.

7. Olkoon $\mathcal{R} = \{S_1, S_2, \ldots S_n\}$ joukko merkkijonoja vakioaakkostossa. Jonon $S_i$ *lyhin yksilöivä osajono* on lyhin jono, joka esiintyy $S_i$:ssä mutta ei muissa $\mathcal{R}$:n jonoissa. Hahmottele lineaarisessa ajassa toimiva algoritmi, joka löytää kaikkien $\mathcal{R}$:n jonojen lyhimmän yksilöivän osajonon.

8. Muodosta jonon `mississippi` loppuosataulukko käyttäen luennoilla esitettyä osarekursiivista algoritmia.

9. Minkä jonon Burrows-Wheeler-muunnos on `RTTRRAA$II` ?