Merkkijonomenetelmät (kevät 2007)

Harjoitus 2

Jarkko Toivonen pe 2.2. klo 10.15–12 CK107

1. Triviaali- ja KMP-algoritmit ovat *aakkostoriippumattomia*: ne tekevät vain merkkivertailuja ja niiden toiminta ei riipu aakkostosta, erityisesti niiden pahimman tapauksen aikavaatimus ei riipu aakkoston koosta. BM-algoritmin toiminta sen sijaan riippuu aakkostosta.

   (a) Anna esimerkkejä tilanteista, joissa aakkostoriippuvuus rajoittaa BM-algoritmin käyttökelpoisuutta.

   (b) Saadaanko BM-algoritmista aakkostoriippumaton, jos toinen sen heuristiikoista poistetaan?

   (c) Ovatko Karp-Rabin-algoritmi ja shift-or-algoritmi aakkostoriippumattomia? Onko jollakin niistä aakkostoriippuvuus erilaista eli enemmän tai vähemmän rajoittavaa kuin BM-algoritmilla?

2. Hahmottele seuraavien algoritmien yleistys monen hahmon tapaukseen:

   (a) Karp-Rabin

   (b) Shift-or

   Aikavaatimuksen tulee olla parempi kuin kunkin hahmon etsiminen erikseen saman algoritmin perusversiolla.

3. Muodosta AC-algoritmin mukainen jononsovitusautomaatti hahmojoukolle {`SAKARI`, `SAKU`, `ARI`, `MARKO`, `ILMARI`}.

4. Kehitä menetelmä, jolla NFA, joka säännölliselle lausekkeelle A löytää $L(A)$:n esiintymät tekstistä S, pystyy tulostamaan myös ne $L(A)$:n jonot, jotka ko. paikassa esiintyvät. Esim. hahmolla $P = b(\texttt{ac} \cup \texttt{b})^*\texttt{ba}$ tekstistä $S = \texttt{ccbbbacacbacc}$ tulisi löytää 5 esiintymää: $S[2\ldots5]$, $S[3\ldots5]$, $S[2\ldots10]$, $S[3\ldots10]$ ja $S[4\ldots10]$.

5. Kuten tehtävässä 1 mainittiin, triviaali- ja KMP-algoritmit tekevät vain merkkivertailuja. Voitaisiinko niitä käyttää hahmonsovitukseen jokerimerkeillä käyttämällä sopivaa vertailuoperaattoria merkkien vertailuun? Perustele.

6. Laske jonojen `BANAANI` ja `ANANAS` editointietäisyys.