Merkkijonomenetelmät (syksy 2003)

Harjoitus 3

Ryhmä	aika	paikka
Veli Mäkinen	to 9.10 klo 8.30–10	A319
Anna Pienimäki	pe 10.10 klo 8.30–10	B453

1. Simuloi Shift-or-algoritmia, kun teksti on `TOTATOTOA` ja hahmo on `TOTO`.

2. Oletetaan, että hahmossa yksittäisen merkin paikalla voi olla myös jokerimerkki tai merkkiluokka. Jokerimerkin ? kohdalla sallitaan hahmon esiintymässä mikä tahansa merkki. Merkkiluokan (esim. [`a`, `f-h`]) kohdalla sallitaan hahmon esiintymässä mikä tahansa tähän merkkiluokkaan kuluva merkki. Muunna Shift-or-algoritmia niin, että se löytää yleistetyn hahmon esiintymät.

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

4. Todista, että Algoritmi 1.6.4 (Aho-Corasick-esiprosessointi II) laskee $fail$-funktion ajassa $O(m)$. (Oletetaan, että aakkoston koko on vakio, ja funktiosta *output* ei tarvitse välittää.)

5. a) Muodosta NFA säännölliselle lausekkeelle $(\mathtt{a} + \mathtt{b})(\mathtt{ac} + \mathtt{b})^*\mathtt{ba}$ Thompsonin konstruktion (Lause 1.7.4, Kuva 1.2) avulla.

   b) 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. tekstistä $S =$ `ccbbbacacbacc` menetelmän tulisi löytää a) kohdan säännölliselle lausekkeelle 5 esiintymää: $S_{3...6}$, $S_{4...6}$, $S_{3...11}$, $S_{4...11}$, ja $S_{5...11}$.