Merkkijonomenetelmät (kevät 2005)

Harjoitus 3

Pasi Rastas ma 21.2 klo 10.15–12 C221

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

2. 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 $P = b(\mathtt{ac} \cup \mathtt{b})^*\mathtt{ba}$ tekstistä $S = \mathtt{ccbbbacacbacc}$ tulisi löytää 5 esiintymää: $S[2\ldots5]$, $S[3\ldots5]$, $S[2\ldots10]$, $S[3\ldots10]$ ja $S[4\ldots10]$.

3. Modifioi shift-or-algoritmia siten, että se toimii myös, jos hahmo ja teksti sisältävät jokerimerkkejä.

4. Määritellään *occur*-vektorin yleistys. Olkoot $A, B \subseteq \Sigma$ merkki*joukkoja* ja $U, V \in \Sigma^*$ jonoja.

   - $occur(A,U)[i] = 1$, jos $U[i] \in A$, ja 0 muuten.
   - $OCCUR_{A,B}(U,V) = occur(A,U) \diamond occur(B,V)$.

   Luentojen kaavan (1.1) avulla $\circ$-operaatio voidaan laskea käyttäen $\mathcal{O}(\sigma^2)$ $\diamond$-operaatiota. Anna uusi, *occur*-vektorin yleistykseen perustuva kaava, joka tarvitsee vain $\mathcal{O}(\log\sigma)$ $\diamond$-operaatiota. (Vihje: Koodaa merkit bittijonoina ja jaa merkit kahteen joukkoon A_i ja B_i positiossa i olevan bitin perusteella.)

5. Näytä, miten luentojen Bird-Baker-algoritmi löytää hahmon

```
A A A B
A B A B
A B B A
A A A B
```

   kaikki esiintymät tekstistä

```
A A A B A B B A
A B A B A A A A
A B B A A B A B
A A A B A B B A
A B A B A A A B
A B B A A B A B
A A A B A B B A
B A B A A A A B
```