Merkkijonomenetelmät (syksy 2003)

Harjoitus 8

Ryhmä	aika	paikka
Veli Mäkinen	to 13.11 klo 8.30–10	A319
Anna Pienimäki	pe 14.11 klo 8.30–10	B453

1. Kappaleen 2.7.1 seulontamenetelmässä hahmon palasia etsitään Aho-Corassic automaatilla tekstistä.

   a) Oletetaan, että hahmon $P = p_1 p_2 \cdots p_m$ palanen $p_{i'} p_{i'+1} \cdots p_i$ löytyy tekstin kohdasta j (siis $p_{i'} p_{i'+1} \cdots p_i = s_{j-(i-i')} s_{j-(i-i')+1} \cdots s_j$). Mikä alue (indeksiväli) tekstistä pitää tarkastaa dynaamisella ohjelmoinnilla kun sallittu virheiden määrä on k?

   b) Eri palasten esiintymiin liittyvät tarkistettavat alueet voivat leikata toisiaan. Tällöin kannattaa laskea näiden alueiden yhdiste ja suorittaa tarkistusalgoritmi kerralla koko yhdisteen alueella. Miten alueiden yhdiste lasketaan lineaarisessa ajassa?

2. Galil-Giancarlo-algoritmissa käytetään taulukkoa $prefix[1 \ldots m, 1 \ldots m]$, missä $prefix[i', i]$ kertoo hahmon P loppuosien $P_{i' \ldots m}$ ja $P_{i \ldots m}$ pisimmän yhteisen alkuosan pituuden. Esimerkiksi hahmolla $P = \underline{\texttt{AB}}\texttt{C}\underline{\texttt{AB}}\texttt{D}$, $prefix[1, 4] = 2$. Triviaalialgoritmi $prefix$-taulukon täyttämiseksi vie pahimmassa tapauksessa ajan $O(m^3)$. Näytä miten taulukko voidaan täyttää ajassa $O(m^2)$.

3. Havainnollista Galil-Giancarlo-algoritmin toimintaa, kun teksti on `abaabaaaa`, hahmo on `abaaaa`, ja virheiden määrän yläraja on $k = 1$.

4. Löydä $n + 1$ merkin pituinen merkkijono S (s.e. $s_{n+1} = \$$) jonka loppuosatrien koko on $\Theta(n^2)$. Mikä on vastaavan loppuosa*puun* koko?