Merkkijonomenetelmät (kevät 2007)

Harjoitus 1

Jarkko Toivonen   pe 26.1 klo 10.15–12   CK107

1. Luennolla esitetty KMP-algoritmi on itseasiassa Morris-Pratt-algoritmi. Varsinainen Knuth-Morris-Pratt-algoritmi eroaa siitä vain korjaussiirtymien osalta. Se käyttää seuraavaa $fail'$-taulukkoa:

   - $fail'[i] = k$, kun k on suurin luku, jolla $P[0 \ldots k) = P[i-k \ldots i)$ **ja** $P_k \neq P_i$.

   (a) Kuvaa, miten ja miksi $fail'$-taulukko eroaa $fail$-taulukosta.

   (b) Laske $fail$- ja $fail'$-taulukot hahmolle `IKONITIKIKITIKO` ja totea erot.

2. Modifioi luennolla esitettyä KMP-algoritmin esiprosessointia (Algoritmi 1.1.3) laskemaan tehtävän 1 $fail'$-taulukko.

3. BM-algoritmin molemmat heuristiikat antavat pienimmän siirtymän, joka täyttää tietyn ehdon. On myös mahdollista määritellä yhdistelmäheuristiikka, joka antaa pienimmän siirtymän, joka toteuttaa molemmat ehdot.

   (a) Anna esimerkki tilanteesta, jossa yhdistelmäheuristiikka antaa paljon pidemmän siirtymän kuin luentojen algoritmi (eli maksimi kahden perusheuristiikan siirtymästä).

   (b) Onko yhdistelmäheuristiikassa mitään haittapuolia?

4. Olkoon $\Sigma = \{\mathtt{A}, \mathtt{B}\}$, ja numeroidaan aakkoston Σ merkit seuraavasti:

   $$\mathtt{A} \cdots 0$$
   $$\mathtt{B} \cdots 1$$

   Olkoon $q = 7$. Simuloi Karp-Rabin-algoritmin (Algoritmi 1.4.1) toimintaa hahmolla $P = \mathtt{BAAA}$ ja tekstillä $S = \mathtt{ABAABAAABAABA}$. Montako oikeaa/väärää täsmäystä löytyy? Montako merkkivertailua tehdään?

5. Merkkijonot x ja y ovat toistensa *rotaatioita*, jos on olemassa jonot u ja v, joilla $x = uv$ ja $y = vu$. Esimerkiksi *abcde* ja *deabc* ovat toistensa rotaatioita. Hahmottele *lineaarisessa* ajassa toimiva algoritmi, joka tutkii, ovatko annetut jonot x ja y toistensa rotaatioita. (Vihje: käytä lineaarisessa ajassa toimivaa hahmonsovitusalgoritmia (esim. KMP) alirutiinina.)