

Määrittelydokumentti

Teen pakkausohjelman Huffmanin koodilla, mikä käyttää aputietorakenteena toisiinsa linkitetystä solmuista muodostuvaa binääripuuta, taulukkoa, mihin puut talletetaan, sekä taulukkoa, mihin kirjainten koodatut binääriesitykset talletetaan.

Ohjelma pakkaa teksti- tai binääritiedoston pienempään tilaan käyttämällä Huffmanin koodia. Ohjelma saa syötteenä tiedoston, teksti- tai binääritiedoston, jossa olevien merkkien esiintymistiheys lasketaan. Merkit esiintymistiheyksineen talletetaan taulukkoon nousevaan järjestykseen. Tähän tarkoitukseen käytetään solmuja, jotka talletetaan taulukkoon. Solmuista muodostetaan binääripuita, mitkä talletetaan samaan tauluun. Lopulta solmuista on muodostettu yksi puu.

Puuta läpikäymällä saadaan toiseen taulukkoon tiedot merkkien uusista binääriesityksistä. Vasemmalle mennessä tulee nolla ja oikealle mennessä yksi. Puun lehtiin on tallennettu merkit. Lehteen päätyessä on saatu selville uusi koodaus kyseiselle merkillä. Useammin esiintyvillä merkeillä on lyhyempi binääriesitys kun harvemmin esiintyvillä.

Koodin muodostuksen aikavaativuus tulee olla $O(n \log n)$, missä n on eri merkkien määrä, ja merkit yhdistellään binääripuuksi. Tiedoston pakkaus- ja purku $O(m * n)$, missä m on pakattavien tai purettavien merkkien määrä ja n on kaikkien merkkien määrä, mihin niitä verrataan.

Lähteet:

<http://www.cprogramming.com/tutorial/computersciencetheory/huffman.html>

<http://www.compressconsult.com/huffman/#conventions>

http://fi.wikipedia.org/wiki/Huffmanin_koodaus