

Viikkoraportti 3

Kolmannella viikolla on ollut muista kiireistä johtuen erittäin niukasti aikaa tiralabralle. Olen toteuttanut Huffman-koodauksen pääkomponentit. Merkkijonojen pakkaus toimii periaatteessa, vaikka se on erittäin hiomaton ja hieman buginen. Tällä hetkellä ohjelma muodostaa Huffman-puun, jonka avulla muodostettavaa kooditaulua käytetään sekä pakkaamiseen että purkamiseen. Kooditaulun tallennus merkkijonomuotoon on toteutettu alustavasti vaikkakin varsin kömpelösti. Tällä hetkellä pakkauksen toteutus käyttää purkamiseen täsmälleen samaa kooditaulua, joka on pakkaamisen yhteydessä on muodostettu, eikä sitä siis ole sisällytetty pakattuun dataan.

Ohjelman toiminta etenee seuraavasti: alustustoimenpiteinä muodostetaan taulukko pakattavan aineiston merkkien esiintymistiheydestä (muodostaTiheystaulu). Sen jälkeen metodi muodostaHuffmanPuu rakentaa tiheystaulun perusteella Huffman-puun, jonka pohjalta metodi muodostaKooditaulu rakentaa kunkin merkin koodaamiseen käytettävän binääriesityksen merkkijonomuodossa.

Itse pakkaus käsittää seuraavat vaiheet. Metodi koodaaBittijonoksi korvaa jokaisen alkuperäisessä merkkijonossa olevan merkin sitä vastaavalla binääriesityksellä, joka haetaan aiemmin luodusta kooditaulusta. Tämän jälkeen metodi bititMerkkijonoksi tiivistää binäärimerkkijonon muodostamalla jokaisesta kahdeksan bitin jaksosta yhden merkin. Näin saatu merkkijono on pakkauksen lopputuote.

Purkaminen aloitetaan metodilla merkitBittijonoksi, joka palauttaa pakatun merkkijonon ykkösiä ja nollia sisältäväksi merkkijonoksi. palautaMerkkijonoksi –metodi muodostaa tämän binääriesityksen ja kooditaulun perusteella merkkijonon, joka on sama kuin alkuperäinen pakkaamaton merkkijono.

Lyhyillä merkkijonoilla päästään tällä hetkellä noin 50 %:n pakkaustehoon, mutta tämä ei sisällä kooditaulua, joka on välttämätön pakatun tiedon purkamiseen.

Seuraavalle viikolle jää Huffman-koodauksen viimeistely ja tiedostonkäsittelyn yhdistäminen ohjelmaan siten, että se lukee syötteen tiedostosta ja kirjoittaa pakatun datan toiseen tiedostoon. Ohjelmassa seuraavana myös omien tietorakenteiden kehittäminen valmiiden toteutusten tilalle, sekä testauksen laajentaminen.