TIETORAKENTEET JA ALGORITMIT -HARJOITUSTYÖ: TIEDON TIIVISTYS VIIKKORAPORTTI I

VILLE TENHUNEN 013618793 22.12.2013

Muutaman päivän työskentelyn jälkeen olen saanut jonkin verran koodia aikaiseksi. Tosin se on vielä hieman sekavassa tilassa, koska aikaa oli rajallisesti. Lähdin ripeästi toteuttamaan Huffmanin koodausta, josta jonkinlaisen alustavan version kerkesin saada kasaan. Koska aikaa tuli käytettyä tähän, jäi teoriaan perehtyminen hieman pienemmälle huomiolle, ja toisaalta teoria oli sen verran vaikeaselkoista, että se siksi jäi toissijaiseksi, kun koodauksen sai käyntiin ensimmäiseen algoritmiin perehtymällä. Tämä näkyy määrittelydokumentissa, jossa en ole osannut selostaa tarkemmin muuta kuin Huffmanin algoritmia. Täydennän sitä lisää kun saan paremman näkemyksen muista osista.

Huffmanin algoritmissa sain muodostettua minimikeon ja sen avulla puun, ja mielestäni myös tuon puun avulla saan järkevästi purettua koodauksen. Sitä en tiedä, että jos koodauksen tekee esiintymistiheyksien perusteella, kuinka saan tallennettua myös tämän puun tiedot. Ilmeisesti jos on tiedossa jokin merkkien todennäköisyysjakauma, sen tiedon voi ottaa jostain annettuna, mutta jos sen puun omin päin koodaa, pitäisi kai sen tiedot myös tallentaa johonkin, jotta koodin saa purettua. Lisäksi en tiedä, miten pitäisi tallentaa luomani binäärikoodi. Nyt olen tallentanut sen vain tekstitiedostona, joka on tietenkin isompi tiedosto kuin alkuperäinen tekstitiedosto.

Tiedon pakkaamisessa onnistuin tekemään toimivan algoritmin, mutta en varmastikaan sellaista kuin oli tarkoitettu. En keksinyt, miten koodausvaiheessa kun puu on luotu saisin puusta koodit kirjaimen perusteella. Päädyin sitten muutamissa kohdissa tehottomampiin ratkaisuihin ja laitoin solmut listaan, jotka kävin läpi etsien oikeaa kirjainta.

Yksikkötestausta en vielä kerinnyt aloittaa, mutta olen testannut koodin osia main-metodissa. Ja mielestäni kaikki toimii virheettömästi. Täydennän myöhemmin yksikkötestauksen ohjelmaan.

Seuraavaksi olisi tarkoitus jatkaa tässä vielä kesken jääneitä asioita, ja aloittaa perehtymään LZW algoritmiin.