

# **TIETORAKENTEET JA ALGORITMIT -HARJOITUSTYÖ: TIEDON TIIVISTYS VIIKKORAPORTTI II**

VILLE TENHUNEN  
013618793  
5.1.2014

Viimeisen parin viikon aikana olen tehnyt LZW algoritmin, aputietorakenteita algoritmeille sekä testejä kaikille luokille. Paljon aikaa on mennyt tiedon hakuun ja materiaaleihin perehtymiseen.

LZW algoritmi on vielä toteutettuna ArrayListin avulla, joten se pitää vielä muokata eri muotoon. Siinä onnistuin keräämään kirjaston indeksit halutusti ja palauttamaan kirjaston pakkauksen purkuvaiheessa. Ongelma on edelleen siinä, kuinka tallennan tiedoston. Tällä hetkellä ideanani on tallentaa indeksit merkkijonona välilyönnein tai pilkuin eroteltuna, mutta tällöin tietysti merkkijonon pituus kasvaa merkittävästi ja hyödyt pienenevät. Tosin jos LZW:n jälkeen ajaisi vielä Huffmanin algoritmin syötteelle, josta suuri osa on pilkkuja tai välilyöntejä, niin tämä ongelma pienenee.

Toteutin ohjelmaan myös hajautustaulun, mutta en ole vielä kerinnyt sitä ottaa käyttöön ohjelmassa muissa luokissa. En ole myöskään vielä testannut pakkausalgoritmien tehokkuutta erilaisilla teksteillä, koska ohjelmoitavaa on vielä tekemättä. Muutamilla lyhyemmillä teksteillä olen algoritmeja testannut, ja jos onnistun haluamallani tavalla Huffmanin algoritmin tallentamaan, niin ilman 'puun' tallemista koko on pienentynyt alle puoleen. Puun tallennusta en ole vielä edes loppuun asti pohtinut. Lisäsin ohjelmaan englannin ja suomen kielen kirjainten todennäköisyystaulukot, mutta niitä käyttämällä koodi jäi 85-95 prosenttiseksi alkuperäisestä. Tähän lienee syynä se, että lähteessäni wikipediassa oli vain kirjainten todennäköisyydet, ja tietysti tarvitsen tiedon myös välimerkkien, erityisesti välilyönnin, pisteen ja pilkun todennäköisyyksistä, koska niitä on merkittävän paljon. Paljon on siis vielä tehtävää.

Viime viikkoraporttia tehdessäni ongelmanani oli muun muassa se, että kuinka saan Huffmanissa luomani Stringiin tallennetun uuden bit-tijonon tallennettua tiedostoon bitteinä merkkijonon sijaan. Siihen omaan nyt suunnitelman, jota en ole vielä ehtinyt toteuttaa. Jos se ei onnistu suunnitellusti, kysyn sitten neuvoa vielä siihen asiaan.