

Mitä olen tehnyt tällä viikolla / Miten ohjelma on edistynyt?

Tällä viikolla olen laiskotellut 5 päivää ja tehnyt kaikki hommat putkeen 2 viimeisen päivän aikana ja tästä syystä jeesus-teippiä paloi taas runsaat määrät. Sain tehtyä binääritiedoston lukemisen, LZW ja Huffmanin koodauksen toimimaan sekä pakatessa että purettaessa. Lisäksi tein ArrayListin siinä määrin kun sitä vaaditaan työssäni. Refactoroin myöskin pari metodia luokkaan jonka vastuulla on Huffmanin puun luonti sekä purettaessa että pakatessa. Huomasin noin 21.00 että LZW EI TOIMI isoilla tiedostoilla, ja koska pitäisi vielä kasa JUnit testejä tehdä en ehdi debuggamaan mitä siinä kusee. Jää ensi viikolla. Niin ja iso tiedosto mitä testasin oli King James Bible, huffman koodaus toimii hyvin tälläkin. Tulokset on projektissa dist kansiossa, pakkasin ja purin sekä readme tiedostoa (mitä viime viikollakin koitin) että raamattua molemmilla. Lisäksi testasin LZW toimintaa myös omatekoisella tekstitiedostolla mihin oli kopioitu wikipedia artikkeli, näytti toimivan OK. Niin ja raamattua pakatessa meni LZWllä todella pitkä aika...

Ja nyt mielenkiintoinen osio raportista eli algoritmien pakkaustulokset!

Tiedosto1 README.TXT vei alunperin 1329 tavua, Huffmanilla pakatessa se vei 1058 tavua ja LZWllä 1012 tavua. 79.61% Huffman ja 76.15% LZW purkuteho prosentteina.

Tiedosto2 bible.txt vei alun perin 5498995 tavua ja huffmanilla onnistui pakkaus(ja purku) onnistuneesti. 3068618 tavua vei pakkauksen jälkeen eli päästiin 55.8% alkuperäisestä koosta ja siedettävän nopeasti meni sekä pakkaus että purku. Olettaisin että toimivalla LZW algoritmilla päästään vielä parempaan tulokseen ja myös huffmanin algoritmissa on vielä parantamisen varaa..

Mitä opin tällä viikolla?

Kuinka Huffmanin koodin purkaus tehdään, kuinka binääritiedosta luetaan, kuinka ArrayList tehdään itse ja myöskin alustavasti tutustuin hashmapin tekemiseen. Opin myös sen että ensi viikolla olisi parasta aloittaa hommat ennen lauantaita ja tehdä niitä tasaisesti koko viikon ajan, ei ollut kovin kiva viikonloppu. Meni kokonaan koneen edessä koodatessa. Toivottavasti en unohda tätä opetusta alkuviiikolla (kuten yleensä käy).

Mikä jäi epäselväksi tai tuotti vaikeuksia?

Asia mihin paloi eniten aikaa oli saada huffmanin koodauksen purku toimimaan oikein. Ja nimenomaan se kirjoitetun puun lukeminen tiedostosta oli hankala juttu. Lopulta kiersin tämän kohdan kirjoittamalla pakatun tiedoston alkuun puun sijasta merkkien esiintyvyydet ja homma lähti nopeasti pyörimään. Vihaan myöskin JUnit testien tekoa ja kaikki noi testit on tehty viimeisen 4 tunnin aikana. Viikon suurin ongelma oli oma laiskuus ja paskuus (niin kuin se on aina).

Mitä teen seuraavaksi?

Seuraavaksi yritän debuggaa mikä sitä LZW algoritmia oikein vaivaa. Aion myös tehdä HashMapin ja PriorityQueueen ite. Aloitan suorituskkyttestauksen ja mahdollisesti teen hiukan koodin refactorointia, näyttää turhan rumalta tällä hetkellä.. Lisäksi pitäisi miettiä mitä muita algoritmeja voisin toteuttaa viimeisen 2 viikon aikana. Tavoitteena myös tehdä jonkunlainen alkeellinen tekstikäyttöliittymä, että voisi nopeammin ajella ohjelmaa, alkaa menee hermot siihen että koko ajan pitää buildaa projektia uusiksi kun algoritmeja vaihtaa. Otan myös tavoitteeksi että tällä viikolla pääsisin käymään laitoksella ohjauksessa. Ja yritän ottaa peukun pois perseestä ja aloittaa hommat viimeistään keskiviikkona!