

# Testausdokumentti

## **Automaattiset Testit:**

Projektin eri tietorakenteiden toimivuutta on testattu erikseen JUnit testejen avulla. Rakenteiden eri osaalueet ovat todettu toimiviksi erilaisilla syöteillä.

Huffmannin algoritmin toimivuus on myös onnistuttu testaamaan JUnit testeillä. Testit jotka on ovat lyhyt kovakoodattu luotu String, 10000 merkin mittainen randomeiden merkkien liuta ja 450kb kokoinen guttenberg.txt joka sisältää englanninkielisen kirjan. Kaikissa instanseissa teksti kompressoituu ja sitten dekompressoituu täsmälleen samaksi tekstiksi kuin alkuperäinen.

Kaksi ensimmäistä testiä käsittelevät liian pieniä tiedostoja niiden kompressoidun muodon ollakseen tehokkaasti pienempi, mutta 450kb kokoisessa tiedostossa koko kutistuu 250kb, eli noin 55% alkuperäisestä.

JUnit testit ovat näkyvillä ja uudelleen ajettavissa automaattisesti.

## **Ongelmia testauksessa:**

Sivuhuomiona random teksti välillä ei mene läpi testeistä, epäilen joidenkin ascii merkkejen epäsopivuutta StringBuilder muuttujaan josta appendaan koko stringin tekstidokumenttiin. Appendaus siis tuottaa välillä joitakin ongelmia joidenkin ascii merkkejen kanssa.

Toinen ongelma on esiintynyt tiettyjen tiedostojen kanssa randomin ”ÿ”-merkin ilmestymisenä joko puretun tekstin alkuun tai loppuun riippumatta tiedoston pituudesta tai merkeistä. Tämän jälkeen mikä tahansa muokkaus alkuperäiseen tiedostoon ei poista tätä merkkiä puretusta tiedostosta (tämän tiedoston tyhjäksi jättäminenkin tuottaa puretussa tiedostossa ÿ merkin). En ole varma onko ongelma algoritmissani, FileReaderissä,

Writerissa vai riippuvainen tavasta jolla luon tekstitiedoston alunperin..... Jos tämä kiinnostaa niin tekstidokumentit guttenberg.txt ja buginen.txt sisältävät täsmälleen saman sisällön, mutta buginen antaa syötteeseen y:n vaikka mitä tekisit tekstille.

Kolmas hauska ongelma johon välillä törmää tiedostosta lukiessa ja siihen kirjoittaessa on jonkin tietyn erikoismerkin esim. ä:n vaihtuminen jokaisessa instanssissa joksikin toiseksi erikoismerkiksi. Koska ongelma ei esiinny kaikissa tiedostoissa, ja en linuxin puolella törmännyt ongelmaan, uskon että ongelma johtuu tekstin muodoista (Onko UTF-8, tai jokin muu) ja windowsin oletusasetuksista. En kuitenkaan saanut ongelmaa selvitettyä lopullisesti.

### **Kompressoinnin Testausta:**

Uncompressed	Compressed	Komp Prosentti
--------------	------------	----------------

Normaalia tekstiä:

15 bytes	40 bytes	266.67%
160 bytes	140 bytes	87.5%
9kb	5.10 kb	56.67%
30kb	16kb	52.32%
300kb	167KB	55.66%
450kb	250kb	55.56%
4.3mb	2.39mb	55.58%

Vain muutamaa merkkiä:

5kb	1.18kb	23.6%
-----	--------	-------

Kaikkia meidän aakkosiamme ja kutakin yhtä monta:

5kb	2.89kb	57.8%
-----	--------	-------

Testauksesta näemme että yletämme tavoitteeseemme 40% ja 60% välille kun syöte on tarpeeksi suuri. Erittäin pienillä syötteillä ei

tiivistys ole tehokasta sillä kompressoidun tiedoston tulee kompressoidun datan lisäksi sisältää puurakenne, puurakenteen koko sekä merkkimäärä tiedostossa.