Testausdokumentti

Aineopintojen harjoitustyö: Tietorakenteet ja algoritmit

Joonas Longi

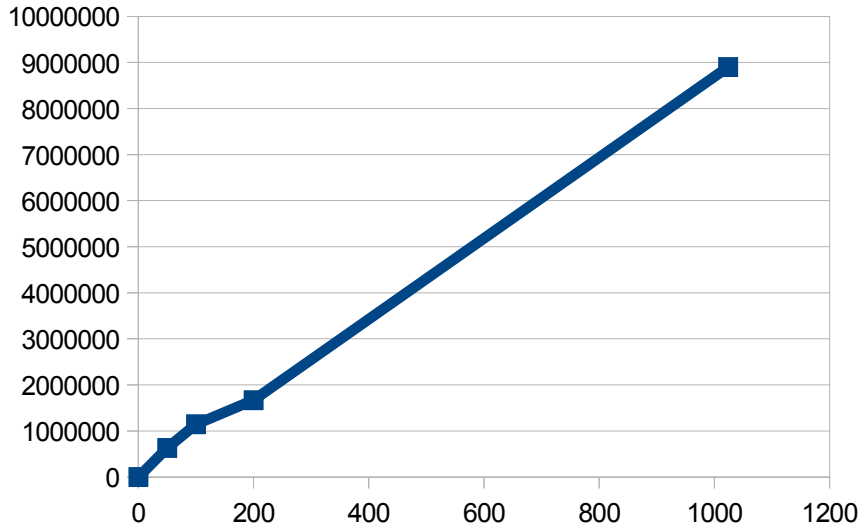
Junit testaus

Ohjelman tärkeimmät metodit ja tietorakenteet on testattu junit testeillä, lyhyillä tekstitiedostoilla, sekä erilaisilla tulosteilla. Lisäksi ohjelmaa on testattu käsin pakkaamalla, purkamalla ja lukemalla ja kirjoittamalla erilaisia tiedostoja. Junit testeissä käytetyt testitiedostot, sekä muut testitiedostot löytyvät ohjelman juurikansiosta.

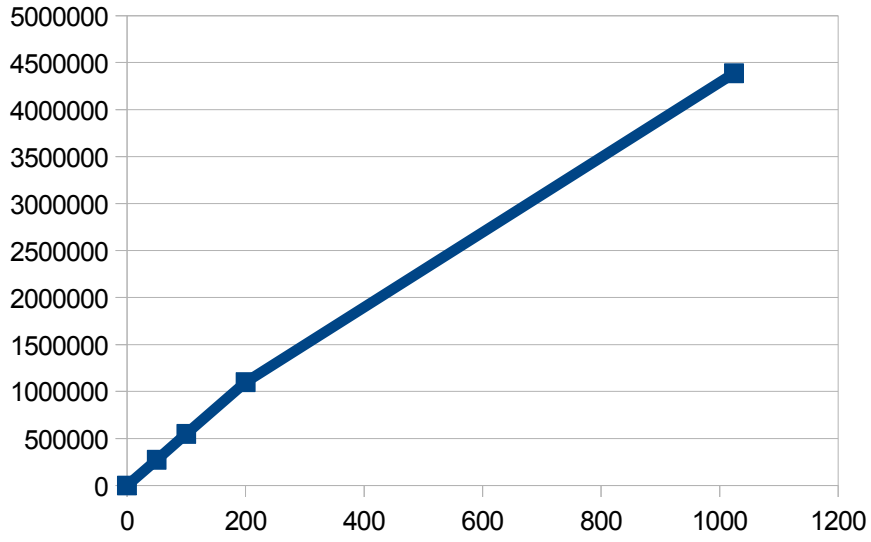
Suorituskykytestaus

Pakkauksen ja purkamisen suorituskykyä on testattu 50MB, 100MB, 200MB ja 1024MB kokoisilla englanninkielisillä tekstitiedostoilla, sekä 50MB, 100MB ja 200MB kokoisilla xml tiedostoilla. Tiedostot ovat pakkauksen ja purkamisen jälkeen todettu samoiksi komentorivin fc tiedosto1 tiedosto2 /B komennolla. Tiedostoina on käytetty http://pizzachili.dcc.uchile.cl/texts.html sivuston tarjoamia tiedostoja. Jokaista pakkausta ja purkamista on kokeiltu 3 kertaa, joista on otettu keskiarvo. Pakattujen tiedostojen koko oli kaikissa testeissä n. 35-70% alkuperäisestä koosta.

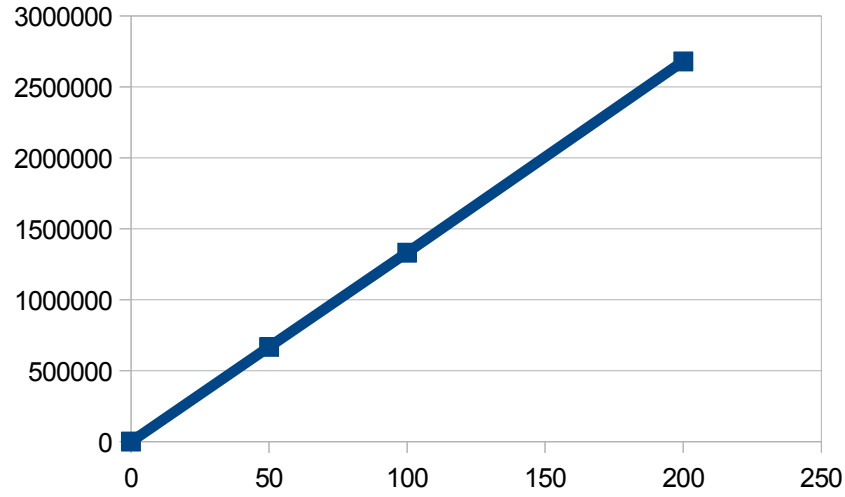
English text pakkaus:



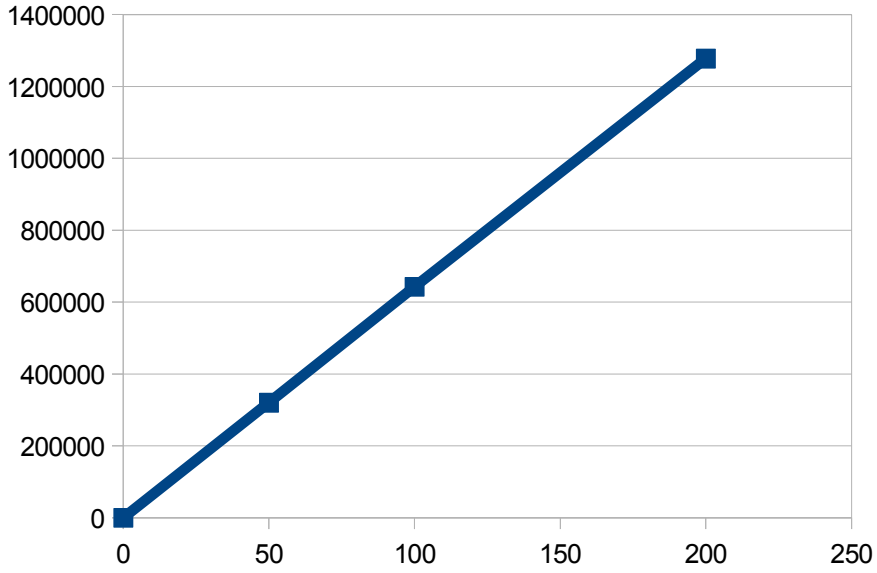
English Text purkaminen



XML pakkaus



XML purku



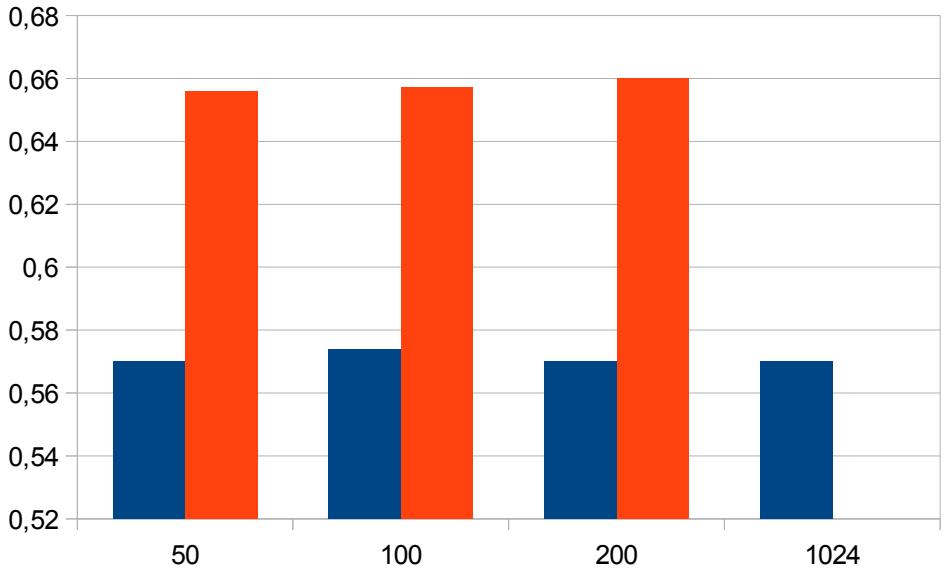
Kuvaajista nähdään, että tavoiteltuun aikavaativuuteen O(n) on päästy. Aikavaativuus on periaatteessa O(n log k), mutta koska k on maksimissaan 256(eri merkkien määrä), on aikavaativuus O(n log 256) = O(n).

Kuvaajista nähdään myös, että xml tiedostot purettiin ja pakattiin hitaammin, kuin englanninkielinen teksti. Tämä johtuu siitä, että xml tiedostot sisältävät paljon erilaisia kirjaimia ja erikoismerkkejä. Huffman toimii tehokkaimmin, kun käytetään mahdollisimman suppeaa aakkostoa.

Aakkoston koko vaikuttaa myös pakatun tiedoston kokoon. Alla on diagrammi pakatun tiedoston koosta suhteessa alkuperäiseen. Huomataan että xml tiedosto, joka sisältää enemmän eri kirjaimia, ei pakkaannu yhtä tiiviiksi, kuin englanninkielinen teksti.

Pakatun tiedoston koko pysyi suhteessa samana xml tiedostoissa (65-66%) ja englanninkielisissä teksteissä (57-58%), koska tekstit sisälsivät melkolailla samat kirjaimet, mutta tuplana tai triplana jne riippuen tiedoston koosta. Eri käsin testatuilla teksteillä on kuitenkin saavutettu parempi pakkaussuhde.

Pakatun tiedoston koko suhteessa alkuperäiseen kokoon
XML
English text



Lähteet

T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein: Introduction to Algorithms,3nd ed., MIT Press, 2009.

Paolo Ferragina, Gonzalo Navarro: Pizza&Chili Corpus, http://pizzachili.dcc.uchile.cl/texts.html. 20.8.2013.