

# Toteutusdokumentti

## Yleisrakenne

Ohjelma koostuu pakkauksista charMap, huffman, main, minHeap, sekä textInterface. CharMap-pakkaus sisältää luokan CharMap, joka on oma toteutus HashMapista, tosin sillä muunnoksella että CharMapista ei pysty hakemaan avaimia yhtä nopeasti kun HashMapissa koska avaimiin ei ole lisätty hashia. Se sisältää myös luokan CharMapEntry, joka toimii CharMapin alkioluokkana. Huffman sisältää luokan Huffman joka huolehtii Huffmanin algoritmin toteutuksesta, sekä apuluokka Node jota käydetään huffman-puun rakentamisessa. Main sisältää luokat TextCompressor joka käynnistää tekstikäyttöliittymän sekä FileIO joka hoitaa tiedostoista lukeminen ja tiedostoihin kirjoittamisen. MinHeap sisältää luokan MinHeap, joka on oma toteutus minimikeosta. TextInterface sisältää luokan TextInterface, joka huolehtii tekstikäyttöliittymästä.

## Algoritmi

Ohjelma käyttää siis tekstin pakkauksessa apunaan Huffmanin algoritmia. Huffmanin algoritmi alkaa laskemalla aineistosta kaikkien eri merkkien lukumäärät. Tämän jälkeen luodaan niin sanottu Huffman-puu, johon sijoitetaan sekä merkkejä että apusolmuja. Kun puu on luotu, kaikille merkeille luodaan puun avulla oma binääriesitys, eli jono nollia ja ykkösiä. Puun rakenne on sellainen, että minkään merkin "koodi" ei voi olla toisen merkin alkuosa, joten ei tule koskaan tilannetta vastaan jossa pitäisi jotenkin tietää että käytetäänkö tätä merkkiä vai jatkentaanko etsintää. Tätä varten tarvitaan niitä apusolmuja.

## Aikavaativuus

Algoritmi käyttää apunaan minimikekoa, joten aikavaativuus on teoriassa  $O(n \log n)$

## Puutteet ja parannusehdotukset

Ohjelmassa on useita puutteita. Suurin puute on se, että pakattua tekstiä ei tallenneta yhteen tiedostoon, vaan siihen käytetään yhtä header-tiedostoa johon huffman-puun tiedot tallennetaan, ja yhtä content-tiedostoa, johon itse pakattu tieto tallennetaan. Toinen suuri puute on että ohjelma ei kykene säilyttämään rivinvaihtoja, eli purun jälkeen koko teksti on yhdellä rivillä. Muuten on varmaan vielä joitain bugeja mitä en ole itse ehtinyt löytämään.

## Lähteet

[https://en.wikipedia.org/wiki/Huffman\\_coding](https://en.wikipedia.org/wiki/Huffman_coding)

<http://web.stanford.edu/class/archive/cs/cs106x/cs106x.1174/assnFiles/assign6/huffman-spec.html>