

# Määrittely

## Tiedon tiivistys

Tarkoituksena tuottaa ohjelma, jolle annetaan tiedosto jonka se tiivistää vähemmän tallennustilaa vieväksi jonka se myös kykenee purkamaan takaisin alkuperäiseksi tiedostoksi.

## Algoritmi

Käytän työssäni Huffmanin koodausta. Algoritmi valikoitui useista syistä. Se on ensinnäkin helpompi ymmärtää ja toteuttaa kuin osa pakkausalgoritmeista. Se on myös hyvin yleiskäyttöinen. Lisäksi Huffmanin koodauksella on historiallista merkitystä pitkältä aikaväliltä ja sitä käytetään monien pakkausohjelmien osana yhä tänä päivänäkin.

Toteutan Huffmanin koodauksen normaaliversion sen ollessa adaptiivista Huffmanin koodausta nopeampi algoritmi, vaikka itse pakkaus jääkin hieman vähäisemmäksi. (Jos jää aikaa toteutan molemmat keskenään vertailtaviksi).

## Ohjelma

Ohjelma ottaa syötteenään tiedoston. Tiedosto luetaan tavu kerrallaan. Tavut muunnetaan biteiksi ja käsitellään. Algoritmi määrittelee tiedostossa esiintyville tavuille uudet bittikoodit kunkin tavun esiintymistaajuuden mukaan. Mitä useammin esiintyvä tavu setä vähemmän bittejä sen uudelleen tallentamiseen käytetään. Tiivistetty bittijono muunnetaan tavuiksi ja kirjoitetaan tiedostoon käyttäjän määrittelemään sijaintiin. Tiedoston alkuun talletetaan käytetyt merkit ja niiden frekvenssit.

Ohjelma toimii komentorivillä ja tekstikäyttöliittymässä.

## Aika- ja tilavaativuus

Tavoiteltu aikavaativuus on  $O(n \cdot \log(n))$ .

Tilavaativuus  $O(n * m)$ , missä  $n$  on syötteen koko ja  $m$  erilaisten merkkien määrä syötteessä.