

Toteutusdokumentti

Topi Sarkkinen

013800255

Huffman-koodaus

Ohjelman yleisrakenne

Ohjelman pääluokka hallitsee Huffman-koodauksen toteutusta. Sillä on käytössään Reader- ja Writer-luokat, jotka lukevat ja kirjottavat tiedostoihin. Se myös luo puun Tree-luokasta, jonka rakenteissa käytetään Node-luokkaa.

Ohjelmassa on myös käytössä PriorityQ-luokka, joka on prioriteettijono/minimikeko. Se laittaa puun solmut järjestykseen niiden frekvenssin mukaan, ja pitää ne järjestyksessä heapifyn avulla uusia solmuja lisättäessä.

Linked-luokka toteuttaa yhdensuuntaisen linkitetyn listan, johon talletetaan bittejä.

Aikavaativuus

Pakkaamisen ja purkamisen aikavaativuus on koodin ja empiirisen tutkailun perusteella $O(n \log n)$.

Koko operaation suorittaminen 8500-rivisellä tekstitiedostolla kesti n. 6 sekuntia ja 17000-rivisellä n. 13 sekuntia.

Puutteet ja parannusehdotukset

Vakava puute työssä on se, että erikoismerkit hajottavat sen. Ajanpuutteen vuoksi päätin keskittyä muuhun kuin merkistöongelmien ratkomiseen, joten tällä hetkellä sitä ei voi käyttää ääkkösten tai muidenkaan erikoismerkkien kanssa. Esimerkiksi Gutenbergistä kopioidut kirjat sisältävät yleensä sellaisia lainausmerkkejä, joista ohjelma ei pidä. Projektikansiossa on testi.txt-niminen tiedosto, johon on kopioitu matemaatikoilta saatu hämärä sähköposti ilman ääkkösiä muutama tuhat kertaa. Olen käyttänyt sen variaatioita testauksessa.

Puute on myös se, että ohjelma ei vielä käytännössä ole kauhean hyödyllinen, sillä tiedostojen purkamiseen käytetään ohjelman sisäistä puuta. TODO: Kirjoita koodatun tiedoston alkuun sanakirja symboleista ja Huffman-koodeista.