

Testausdokumentti

Huffman-koodauksen ja Lempel-Ziv-koodauksen toteutusta on tarkoitus testata käyttämällä syötteenä erilaisia datatiedostoja, kuten tekstiä, kuvaa, ääntä jne. Keskeisin mielenkiinnon kohde tässä testauksessa on pakkauksen tehokkuus, eli se, kuinka paljon pienempään tilaan alkuperäinen data saadaan tiivistettyä. Tämä riippuu luonnollisesti erittäin paljon käytettävän syötteen laadusta. Ensimmäiset testit on tehty tekstitiedostoilla, ja tehtäväkuvauksessa mainittuun 40-60% tiivistykseen pääseminen vaikuttaa olevan hyvin saavutettavissa.

Algoritmien aikavaativuuden tutkiminen ei poikkeuksellisesti ole tiivistämisen osalta kovin mielenkiintoinen tutkimuskohde. Kuten jo määrittelydokumentissa on esitetty, on algoritmien aikavaatimus vakiokokoista kooditaulukkoa ja sanakirjaa käytettäessä $O(n)$. Tehokkuuden testaaminen suoritusajakoja mittaamalla on häiriöaltista, koska toteutuksessa käytetään vahvasti levy-I/O:ta, joka vie luonnollisesti paljon aikaa ja on vaikeasti ennustettavissa/toistettavissa esim. välimuistien vaikutuksen takia.

Tekstitiedoston tiivistys

Aineistona on käytetty Gutenberg-projektista ladattua tekstitiedostoa, joka sisältää Juhani Ahon romaanin Rautatie, joka löytyy ohjelman juurihakemistosta nimellä rautatie.txt. Tiedoston koko on 203 075 tavua.

Tulokset - Huffman-koodaus

Pakkaamiseen kulunut aika: 737 ms, 737 ms, 729 ms

Purkamiseen kulunut aika: 1028 ms, 1142 ms, 980 ms

Pakatun tiedoston koko: 117 760 tavua (58 % alkuperäisestä)

Tulokset - Lempel-Ziv-koodaus

Pakkaamiseen kulunut aika: 3548 ms, 4136 ms, 4100 ms

Purkamiseen kulunut aika: 184 ms, 200 ms, 212 ms

Pakatun tiedoston koko: 97 623 tavua (48 % alkuperäisestä)

Äänitiedoston tiivistys

Aineistona on käytetty miesäänen puhetta sisältävää aaltoäänitiedostoa male.wav. Tiedoston koko on 816 496 tavua.

Tulokset - Huffman-koodaus

Pakkaamiseen kulunut aika: 3353 ms, 3317 ms, 3252 ms

Purkamiseen kulunut aika: 5431 ms, 5594 ms, 5414 ms

Pakatun tiedoston koko: 696 834 tavua (85 % alkuperäisestä)

Tulokset – Lempel-Ziv-koodaus

Pakkaamiseen kulunut aika: 17072 ms, 15345 ms, 16915 ms

Purkamiseen kulunut aika: 947 ms, 1110 ms, 1059 ms

Pakatun tiedoston koko: 904 047 tavua (111 % alkuperäisestä)

Bittikarttakuvatiedoston tiivistys

Aineistona on käytetty ruutukaappaus.bmp –tiedostoa, jonka koko on 3 098 322 tavua.

Tulokset - Huffman-koodaus

Pakkaamiseen kulunut aika: 10393 ms, 10242 ms, 11180 ms

Purkamiseen kulunut aika: 11578 ms, 8201 ms, 8054 ms

Pakatun tiedoston koko: 989 2154 tavua (32 % alkuperäisestä)

Tulokset – Lempel-Ziv-koodaus

Pakkaamiseen kulunut aika: 21229 ms, 20941 ms, 22839 ms

Purkamiseen kulunut aika: 1134 ms, 1220 ms, 1152 ms

Pakatun tiedoston koko: 1 040 379 tavua (34 % alkuperäisestä)