

TIETORAKENTEET JA ALGORITMIT -HARJOITUSTYÖ: TIEDON TIIVISTYS TOTEUTUSDOKUMENTTI

VILLE TENHUNEN
013618793
12.1.2014

Ohjelmassa on toteutettu Huffman-algoritmi sekä LZW-algoritmi tekstitiedostojen pakkaukseen. Huffmannissa käytetään aputietorakenteina solmuista koostuvaa minimikekoa, josta edelleen muodostetaan solmuista koostuva puu. Solmut sisältävät yksittäisen kirjaimen tiedot, ja solmun avaimena käytetään kirjaimen ascii koodia. Huffmannin algoritmin tulisi toimia ajassa $(n \log n)$, jossa n on syötteen pituus. Tällä hetkellä se toimii vielä neliöllisessä ajassa, koska koodia tulisi vielä tehdä tehokkaammaksi.

Tällä hetkellä LZW-algoritmi toimii ajassa $O(n * m)$, missä n on syötteen pituus ja m on kirjaston koko. Kirjaston koko näyttäisi kasvavan suhteessa syötteeseen, joten aikavaativuus on tällä hetkellä neliöllistä. Kirjaston läpi käynnin voisi toteuttaa hajautustaululla, jolloin sieltä alkion haku toimisi keskimäärin vakioajassa, mutta koska tällöin avaimina pitäisi käyttää ascii koodien yhdisteitä pidemmillekin sanoille, ajattelin, että siinä nopeasti tulisi ylivuotoa. Se jäi kuitenkin ajanpuutteessa toteuttamatta, joten tehokkuus jäi alhaisemmaksi.

Ohjelmaan on luotu pieni tekstikäyttöliittymä, mutta se toimii vielä hieman puutteellisesti, koska algoritmit ei vielä toimi täysin halutusti. Työn teko jäi vielä hieman kesken. Huffmannin algoritmista puun tietojen tallennus pakkauskoodin yhteyteen ei vielä toimi, joten näin ollen koodin purkaminen ei ilman pakkauskoodista lainattua puuta onnistu. LZW puolestaan purkaa koodia hetkittäin hieman virheellisesti, eli vaikka teksti onkin pakkauksen ja purkamisen jälkeen lukukelpoista, löytyy sieltä pieniä virheitä. Lisäksi tosiaan algoritmeja pitäisi vielä tehdä tehokkaammiksi, jotta niillä voisi pakata isompia tiedostoja.