

Tietorakenteiden harjoitustyö

Määrittelydokumentti: Adaptiivinen Huffman-koodaus

Tekijä: Timo Sand

Määrittely

Harjoitustyön aiheena on tehdä ohjelma, joka pystyy pakkaamaan ja purkamaan tekstitiedoston. Tähän tarkoitukseen käytetään "Adaptiivista Huffman" -koodausta. Huffman-koodauksessa pakattavaa dataa käsitellään yksi merkki kerrallaan, eli luetaan merkit bittijonona. Kukin merkki koodataan bittijonoksi, joka on sitä lyhyempi, mitä useammin merkki esiintyy datassa. Adaptiivisessa Huffman-koodauksessa luodaan binääripuu ns. "lennosta", eli jokainen luettu merkki päivittää puuta ja täten merkkejen bittijonoja. Purku tapahtuu tässä tapauksessa vastakkaisessa järjestyksessä, eli bittijono käännetään merkiksi ja puuta päivitetään, jotta seuraava merkki voidaan purkaa. Kun uusi merkki luetaan ensimmäistä kertaa puuhun, niin se kirjoitetaan "selkokielisenä" tiedostoon, jotta siitä pystytään rakentamaan puu kunnolla.

Aika- ja Tilavaativuus

Algoritmille tulee tilavaativuudeksi $O(513)$, sillä ilman tasapainoehtoa puu saattaa täytyä vain yhdeltä puolelta. Haluttu aikavaativuus on $O(513)$, ja luultavasti tämä myös toteutuu.

Tietorakenne: Binääripuu

Binääripuu on puu, jonka jokaisella solmulla on 0-2 lasta. Alkioitten etsiminen, lisääminen tapahtuu ajassa $O(n)$. Puussa on solmuja $2n-1$, joten tehtävän merkistön mukaan voidaan sanoa, kun lehtiä on 257, että solmuja on maksimissaan 513. Huffman-puu on aina täysi, eli jokaisella solmulla on 0 tai 2 lasta. Lehtien määrä siis merkkien desimaaliesityksen mukaan 0-255, sekä NYA-solmu, joka siis kuvastaa ei vielä koodattuja merkkejä.

Binääripuun lisäksi käytetään taulukkoa, jossa suorat viitteet puun alkioihin, täten kaikki puuhun liittyvät operaatiot tulevat vakioaikaiseksi. Indeksoidaan alkioitten merkin bittiesityksen mukaan.

Toiminta

Tiedoston purku ja pakkaus suorittavat samoja toimintoja. Kun tiedostoa luetaan, niin koodipuu "rakennetaan" ja tämä jälkeen kirjoitetaan tiedosto uuteen muotoon. Joka kerta kun ohjelma saa tiedoston, se luo tiedostosta varmuuskopion päätteellä ".bac". Jos mahdollisuuksien rajoissa, ohjelma printtaa uuden tiedoston sisällön ruudulle ja kysyy poistetaanko varmuuskopio. Ohjelmaa käynnistetään kutsumalla sitä komentoriviltä, parametrinä se saa tiedoston tai polun tiedostoon, mahdollisesti pystyy ottamaan useamman tiedoston syötteenä. -h tai --help syötettä käyttämällä tulostuu ruudulle ohjelman käyttöopas. Valmistuttuaan ohjelma tulostaa käytetyn ajan ja jos toteutettavissa käytetyn prosessoroitehon ja muistin määrän.