

Määrittelydokumentti

Aineopintojen harjoitustyö: Tietorakenteet ja algoritmit

Joonas Longi

Määrittelydokumentti

•Algoritmit ja tietorakenteet

Toteutan työssäni tiedoston pakkaamisen Huffmanin algoritmilla. Toteutukseen vaadittavat tietorakenteet ovat Huffmanin puu, prioriteettijono, sekä jono. Huffmanin algoritmilla saavutetaan yleensä tavoiteltu tilansäästö 20-90% (Cormen, 2009). Siinä merkeille muodostetaan Huffmankoodit. Mitä yleisempi merkki on, sitä lyhyempi Huffman koodi sille annetaan. Huffmankoodit saadaan muodostettavasta Huffman puusta kulkemalla juuresta lehtiin. Vasemmalle kuljettaessa koodiin lisätään 0 ja oikealle kuljettaessa 1. Puun lehdet sisältävät tiedostossa olevat merkit. (Cormen 2009)

•Ongelma

Ongelmana harjoitustyössä on tiedon pakkaaminen ja tiivistäminen. Tiedosto halutaan pakata niin, että se vie 20-90% vähemmän tilaa. Tiedosto pitää pystyä myös purkamaan alkuperäiseen muotoonsa.

Valitsin kyseisen ongelman, koska se tuntui kiinnostavalta. Tiedoston pakkaaminen on tärkeä ongelma, ja halusin oppia perusteet siitä miten ongelma yleensä ratkaistaan.

•Ohjelman syötteet

Ohjelmalle syötetään tiedoston polku, uuden tiedoston nimi, sekä valitaan hautaanko purkaa vai pakata.

•Aika- ja Tilavaativuudet

Toteutuksessa aikaa vievin osuus on pakattavan tiedoston lukeminen ja kirjoittaminen. Tämän aikavaativuus on $O(n)$, missä n on tiedoston sisältämien merkkien lukumäärä. Varsinaisen Huffmankoodauksen tavoiteltava aikavaativuus on $O(n \log n)$ (Wikipedia, 2013). Tässä n on erilaisten merkkien lukumäärä (max. 256)

•Lähteet

Wikipedia, Huffman-coding, http://en.wikipedia.org/wiki/Huffman_coding (viitattu 1.8.2013)

T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein: Introduction to Algorithms, 3rd ed., MIT Press, 2009.