

Tehtävä 2: Tiedon tiivistäminen

Ratkaisut

Kysymys 1. Teksti abba cd koodataan bittijonoksi

1110011001110001101

Kysymys 2. Bittijono vastaa tekstiä: a bed added

Kysymys 3. Tiedostossa on sata tuhatta merkkiä, joten kiinteänpituisen 3-bittiseen koodaamiseen tarvitaan 300 000 bittiä.

Huffman-koodin viemä tila voidaan laskea seuraavasti:

- Merkin a koodin (0) pituus on 1. Tekstissä on 45 tuhatta merkkiä a, joten a-merkkien koodaamiseen tarvitaan yhteensä 45 tuhatta bittiä.
- Merkin b koodin (101) pituus on 3. Tekstissä on 13 tuhatta merkkiä b, joten b-merkkien koodaamiseen tarvitaan yhteensä 39 tuhatta bittiä.
- Merkin c koodin (100) pituus on 3. Tekstissä on 12 tuhatta merkkiä c, joten c-merkkien koodaamiseen tarvitaan yhteensä 36 tuhatta bittiä.
- Merkin d koodin (111) pituus on 3. Tekstissä on 16 tuhatta merkkiä d, joten d-merkkien koodaamiseen tarvitaan yhteensä 48 tuhatta bittiä.
- Merkin e koodin (1101) pituus on 4. Tekstissä on yhdeksän tuhatta merkkiä e, joten e-merkkien koodaamiseen tarvitaan 36 tuhatta bittiä.
- Merkin f koodin (1100) pituus on 4. Tekstissä on viisi tuhatta merkkiä f, joten f-merkkien koodaamiseen tarvitaan yhteensä 20 tuhatta bittiä.

Laskemalla yhteen eri merkkien koodaamiseen tarvittavat bittimäärät saadaan, että Huffman-koodauksessa tarvitaan yhteensä 224 000 bittiä.

Säästynyt bittimäärä saadaan suorittamalla vähennyslasku

$$300\,000 - 224\,000 = 76\,000.$$

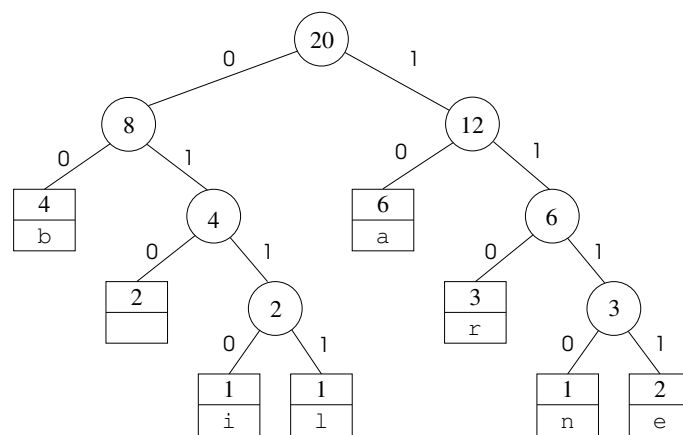
Käyttämällä Huffman-koodausta säästyy siis 76 tuhatta bittiä verrattuna kiinteänpituisen koodaukseen.

Kysymys 4. Muodostetaan ensin tekstissä `a bearable barbarian` esiintyvien merkkien lukumäärätaulukko:

merkki	a	b	e	i	l	n	r	
lukumäärä	6	4	2	1	1	1	3	2

Viimeinen sarake tarkoittaa välilyöntien määrää tekstissä.

Taulukon merkkimääriä vastaava Huffman-puu on:



Huffman-puun perusteella saadaan merkeille seuraavat koodaukset:

merkki	a	b	e	i	l	n	r	
koodi	10	00	1111	0110	0111	1110	110	010

Tekstin `a bearable barbarian` koodaus on siis

1001000111110110100001111111010001011000101100110101110

Huom! Muitakin laillisia tekstin perusteella muodostettuja Huffman-puita ja -koodeja löytyy. Tässä esitetty koodaus ei siis ole ainoa oikea.