# Kiegészítés az órához

Horváth Árpád

2015. október 9.

Az órán adottak voltak az egyes szimbólumok gyakoriságai (darab):

A:2 B:3 C:3 D:25 E:30 F:37

Ezalapján a Huffman-algoritmust végigcsinálva, a következő kódokat rendeltük az egyes jelekhez:

A:11010 B:11011 C:1100 D:111 E:10 F:0

Ebből azt kaptuk, hogyha mind a 100 jelet kódoljuk, ahhoz összesen 209 bit kell, egy jelre tehát $209/100 = 2.09$ bit jut.

Az órán szereplő forrás információelméleti jellemzői octave-val kiszámolva:

```
octave:1> p = [0.02 0.03 0.03 0.25 0.30 0.37]
p =

0.020000   0.030000   0.030000   0.250000   0.300000   0.370000

octave:2> I = -log(p)/log(2)
I =

5.6439   5.0589   5.0589   2.0000   1.7370   1.4344

octave:3> H = sum(p.*I)
H =  1.9682
octave:4> Hmaxn = log(6) / log(2)
Hmaxn =  2.5850
octave:5> eta = H / Hmaxn
eta =  0.76142
octave:6> R = 1-eta
R =  0.23858
```

Az I sorvektorban vannak az egyes jelek információtartamai, pl. a C-é a harmadik: kerekítve 5.06 bit. Összehasonlíthatjuk az információtartalmat a kapott bitsorozatok hosszával. Van ahol egyik, nagyobb, van ahol a másik.

Az entrópia 1.9682 bit/jel. A Huffman-kódban ennél csak több (2.09 bit/jel) átlagos bitszámmal tudtunk kódolni. A Huffman-kóddal nem tudjuk ugyan elérni az entrópia értékét, azt majd csak az aritmetikaival fogjuk tudni elérni. A Huffman-kód mégis optimális (kb. legjobb) egy bizonyos szempontból. Nem lehet ugyanis a Huffmann-kódnál (az órán említett módon előállított kódnál) hatékonyabb olyan kódot készíteni, amelynél minden egyes jelhez adott bitsorozatot rendelünk. Az aritmetikainál majd nem lehet szétdarabolni a bitsorozatot, hogy ettől eddig ehhez a szimbólumhoz tartozik, hanem az egész fogja az üzenetet meghatározni.