

Lutero Simone
4801326

Codifica di Huffman

FILES:

- huffman.cpp: codice per Codificare
- compressed.cpp: codice che comprime il file della codifica
- random.txt: testo generato casualmente
- compressed.txt: testo compresso

$$X = \{x_a, x_b, x_c, x_d, x_e, x_f, x_g\}$$

In x ho M caratteri diversi con $M=7$

$$\text{Freq A} = 50'000 \Rightarrow P_A = \frac{50'000}{100'000} = 1/2$$

$$\text{Freq B} = 20'000 \Rightarrow P_B = 20'000/100'000 = 1/5$$

$$\text{Freq C} = 15'000 \Rightarrow P_C = 15'000/100'000 = 3/20$$

$$\text{Freq D} = 5'000 \Rightarrow P_D = 5'000/100'000 = 1/20$$

$$\text{Freq E} = 5'000 \Rightarrow P_E = 5'000/100'000 = 1/20$$

$$\text{Freq F} = 3'000 \Rightarrow P_F = 3'000/100'000 = 3/100$$

$$\text{Freq G} = 2'000 \Rightarrow P_G = 2'000/100'000 = 1/50$$

$$H(x) = \sum_{i=1}^M P(x_i) \log_2 \frac{1}{P(x_i)} = 1/2 \log_2 2 + 1/5 \log_2 5 + 3/20 \log_2 \frac{20}{3} + 2 \left(\frac{1}{20} \log_2 20 \right) + \frac{3}{100} \log_2 \frac{100}{3} + \frac{1}{50} \log_2 50 \approx 2.07$$

Usando il programma implementato in c++ ho ricavato la codifica di Huffman:

A = 0

B = 10

C = 110

D = 11110

E = 11111

F = 11101

G = 11100

$$\hookrightarrow L(x) = \frac{1}{2} \cdot 1 + \frac{1}{5} \cdot 2 + \frac{3}{20} \cdot 3 + \frac{1}{20} \cdot 5 + \frac{1}{20} \cdot 5 + \frac{3}{100} \cdot 5 + \frac{1}{50} \cdot 5 = \frac{21}{10} = 2.1$$

Visto che $2.1 > 2.07 \rightarrow$ si poteva fare un po' meglio di così