

Laboratori 2 CDI-FIB

Curs 2022/2023

6 de març de 2023

Codificació i descodificació. Es parteix d'una correspondència entre les paraules codi de dos codis prefixos, que poden ser sobre el mateix alfabet o sobre alfabetos diferents, en el format

$$\text{corr} = [(a_1, b_1), (a_2, b_2), \dots, (a_n, b_n)]$$

on les a_i són paraules d'un codi prefix $\mathcal{C}_A \subseteq A^*$ i les b_i són paraules d'un codi prefix $\mathcal{C}_B \subseteq B^*$.

Escriuiu una funció `encode(txta, corr)` que té com a entrada un text $\text{txta} \in \mathcal{C}_A^*$ i retorna el text codificat $\text{txtb} \in \mathcal{C}_B^*$ corresponent, que s'obté substituint cada paraula codi a_i per la paraula codi b_i .

Escriuiu una funció `decode(txb, corr)` que fa la descodificació corresponent. Naturalment, ha de complir-se que

$$\text{decode}(\text{encode}(\text{txta}, \text{corr}), \text{corr}) = \text{txta}$$

per a tot text $\text{txta} \in \mathcal{C}_A^*$.

El cas més important és quan \mathcal{C}_1 és simplement el conjunt de les lletres de l'alfabet A .

Codis canònics. Escriuiu una funció `canonical_code(list, q, alf)` amb paràmetres una llista de nombres enters positius $\text{list} = [\text{long}_1, \text{long}_2, \dots, \text{long}_n]$, un enter $q \geq 2$ i una llista $\text{alf} = [a_1, a_2, \dots, a_q]$ de q lletres d'un alfabet (donades en l'ordre que es vol que sigui l'“ordre alfabètic”). La funció ha de retornar el codi prefix canònic sobre l'alfabet alf format per les paraules amb les longituds de la llista list (o donar un missatge d'error si no existeix cap codi amb aquesta propietat).

Codi de Shannon. Escriuiu una funció `shannon_code(src)` amb paràmetre una llista de la forma

$$\text{src} = [(a_1, w_1), (a_2, w_2), \dots, (a_n, w_n)]$$

on els a_i són lletres d'un alfabet i els w_i són nombres enters positius (“pesos” que es poden interpretar com el nombre de vegades que la lletra a_i surt en un text determinat).

Siguei w la suma de tots els w_i (el nombre total de lletres en el text) i siguin $p_i = w_i/w$ les probabilitats corresponents.

La funció ha de retornar el codi prefix canònic de Shannon amb paraules de longituds $\ell_i = \lceil -\log_2 p_i \rceil$.