

O seguinte boletín de exercicios deberase enviar ao profesor Antonio Fariña á dirección (antonio.lbd@gmail.com) antes do **día 17 de abril de 2012**.

1. Un estudo empírico realizado sobre un texto de 16Gb ( $2^{34}$  bytes) amosou que había 10000 palabras diferentes. Segundo a lei de Heaps, ¿é ese un valor esperado? Xustifica a túa resposta
2. Explica as diferencias básicas entre unha técnica de compresión *fixed-to-variable* e unha *variable-to-fixed*. ¿Como consiguen comprimir? Indica algunha técnica de compresión que clasificarías como *fixed-to-variable* e *variable-to-fixed*.
3. Exercicio.

Dado o texto: CABABAAADAEAAAAAAAB

¿Cal sería a probabilidade coa que se codificaría o 7º símbolo (A) de usar:?

- (a) Un modelado semistático
- (b) Un modelado dinámico

4. Exercicio. Dado o texto: CADABCDEABABCABADABA  
Un esquema de codificación que asigna os códigos:

A → 0  
B → 10  
C → 110  
D → 1110  
E → 11110

Calcúlese:

- (a) A lonxitude media dos códigos resultantes
- (b) O valor da entropía “H” de orde cero ( $H_0$ ). ¿Podería ser este valor maior que a lonxitude media dos códigos obtidos pola codificación? Razona a resposta.
- (c) A redundancia da codificación.

5. Dada a codificación:

A → 0  
B → 10  
C → 1100  
D → 1101  
E → 11100

É unha codificación libre de prefixo? De redundancia mínima? Libre de sufixo?

6. Codifica con Huffman clásico o seguinte texto FFEADBBCDGAABCDAAEAAA  
Amosa:

- (a) a árbore Huffman creada ao aplicar o algoritmo de Huffman básico.
- (b) o número de bits que contera a secuencia final comprimida
- (c) a lonxitude media da codificación obtida
- (d) Unha árbore de Huffman canónica que sexa equivalente á obtida no punto a).

7. Dado o texto

texto 

a	b	r	a		c	a	d	a	b	r	a		p	a	r	a		t	r	a	c	a
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23

Calcúlese a táboa de saltos (array d[]) que usará o algoritmo de horspool cando se busque o patrón “bra”.

Indíquese cantas comparacións “byte a byte” entre os caracteres do patrón e do texto se realizan durante esa búsqueda. ¿e cando o patrón buscado é “otra” en lugar de “bra”?

8. Que dúas leis principais permiten caracterizar o texto en linguaxe natural. Explica brevemente o seu interese se se pretende crear un compresor orientado a palabras (frente a un orientado a caracteres).

9. Asumindo que estivesemos a usar bytes “especiais” de só 3 bits (b=3)...

9.1. indíquese cales serían os códigos que se lles asignarían ás palabras do seguinte vocabulario (palabras A..J coas probabilidades que aparecen na seguinte táboa).

9.2. Indíquese tamén a lonxitude media dos códigos e o valor da Entropía de orde 0.

9.3. Pode ser dito valor da entropía de orde 0 superior á lonxitude media do código do Plain Huffman? Xustifica a túa resposta.

Palabra	Prob	Plain.Huffman	Tagged Huffman	(6,2)-DC	(5,3)-DC	(4,4)- DC	(7,1)- DC
A	0,35	[000]					
B	0,30						
C	0,15						
D	0,05						
E	0,05						
F	0,05						
G	0,02						
H	0,02						
I	0,005						
J	0,005						
Longitud media del código							

10. Indíquese o motivo polo cal os compresores dinámicos estatísticos tradicionalmente deben facer un proceso de compresión simétrico ao que fan os seus descompresores. ¿Coñeces algún compresor estatístico dinámico que non realice un procesamento simétrico ao do correspondente descompresor?