ACTO2 RECUPERACIÓN-SAR

(18/06/2021 - 2 puntos)

Apellidos y Nombre: .	
(IMPORTANT)	🤃 Se deben justificar las respuestas, cálculos redondeados a 3 decimales)

- 1) Se pide dar la secuencia de bits correspondientes a la compresión utilizando codificación variable en bytes de la siguiente postings list: [788, 798, 19755] (0,3 puntos)
- 2) Dada la siguiente lista de símbolos y sus probabilidades se pide:

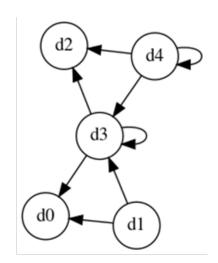
(0,4 puntos)

- a. Construir el árbol de Huffman.
- b. ¿Por qué se consigue compresión con los códigos de Huffman?

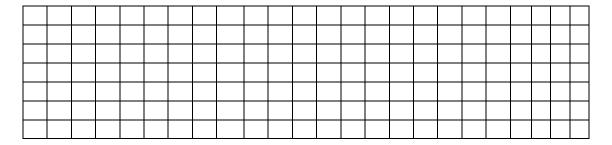
Símbolo	Fr ecu en cia
A	0.30
E	0.20
H	0.08
K	0.02
L	0.03
0	0.18
S	0.19

3) Dadas las siguientes páginas web y los enlaces entre ellas representadas como un grafo, se pide calcular los valores HUB y AUTHORITY de cada página utilizando la aproximación HITS. Realiza 5 iteraciones normalizando al final.

(0,8 puntos)



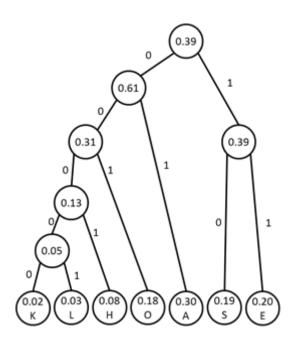
4) Se quiere buscar el patrón "SODA" en el texto "SSOLASLASSEDAS". Para ello se hace una búsqueda aproximada para obtener aquellos segmentos cuya distancia (de Levenshtein) al patrón es menor o igual que 1. Se pide completar la siguiente matriz que corresponde al algoritmo de búsqueda aproximada e indicar las soluciones, es decir, los segmentos del texto que son resultados de la búsqueda. Para el cálculo de la distancia se consideran pesos 1 para las Sustituciones, Inserciones y Borrados. (0,5 puntos)



SOLUCIÓN:

1)
La secuencia de gaps en decimal es: [788, 10, 18957].
La correspondiente secuencia de bits utilizando codificación variable en bytes es: 00000110 10010100 10001010 00000001 00010100 10001101
Es decir: (6, 20+128) (10) (1, 20, 13+128)

2) a



b. La compresión se consigue ya que los símbolos más comunes se codifican con pocos bits y los símbolos menos frecuentes se codifican con códigos más largos.

3)	
	i) la matriz de enlaces

-, -				, ,	
	d0	d1	d2	d3	d4
d0	0	0	0	0	0
d1	1	0	0	1	0
d2	0	0	0	0	0
d3	1	0	1	1	0
d4	0	0	1	1	1

ii) HUB y AUTHORITY utilizando la aproximación HITS. Cinco iteraciones normalizando al final.

			HUB				AUT	HORITY		
	d0	d1	d2	d3	d4	d0	d1	d2	d3	d4
tO	1	1	1	1	1	1	1	1	1	1
t1	0	2	0	3	3	2	0	2	3	1
t2	0	5	0	7	6	5	0	6	8	3
t3	0	13	0	19	17	12	0	13	18	6
t4	0	30	0	43	37	32	0	36	49	17
t5	0	81	0	117	102	73	0	80	110	37
norm	0	0.272	0	0.391	0.337	0.242	0	0.267	0.366	0.125

A	4	3	3	2	2	1	2	3	2	3	3	3	2	1	2		
D	3	2	2	1	1	2	2	2	2	2	2	2	1/	2	2		
0	2	1	1	6	1	2	1	1	2	1	1	1	2	2	1		
S	1	0	0	1	1	1	0	1	1	0	10	1	1	1	0		
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
		S	<u>s</u>	<u>o</u>	Ī	<u>A</u>	S	L	A	S	<u>s</u>	<u>E</u>	<u>D</u>	<u>A</u>	S		

Segmentos encontrados con distancia 1:

"SOLA" de la posición 2 a la 5.

"SEDA" de la posición 10 a la 13.