

Recuperación Acto2 - SAR

(07/06/2019 - 3 puntos)

Apellidos y Nombre:

(NOTA: Se pide justificar las respuestas. Los cálculos se mostrarán redondeados a 2 decimales)

1) Se pide obtener la postingslist a partir de la siguiente secuencia de bits codificada utilizando códigos gamma: **(0,3 puntos)**

111101000101110110111001111100101111000111001

2) Se pide dar la secuencia de bits correspondientes a la compresión utilizando codificación variable en bytes de la siguiente postings list: **(0,3 puntos)**

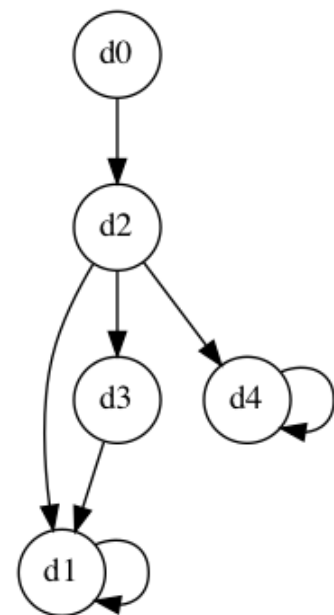
[84, 92, 277, 576]

3) Se pide indicar sobre la tabla, los desplazamientos que se realizarían en una búsqueda por Booyer-Moore del patrón "OFEIO" en la cadena "IWDEIIOEFWOFEIIOCODO". **(0,8 puntos)**

I	W	D	E	I	I	O	E	F	W	O	F	E	I	O	C	O	D				
O	F	E	I	O																	
	O	F	E	I	O																
		O	F	E	I	O															
			O	F	E	I	O														
						O	F	E	I	O											
							O	F	E	I	O										
											O	F	E	I	O						
												O	F	E	I	O					

4) Se pide describir las principales diferencias entre el método HITS y el método PageRank. **(0,6 puntos)**

5) 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 4 iteraciones sin normalización y normaliza la última iteración para que sume 1 (redondea a 2 decimales). **(1 punto)**



1)

Solución:

La secuencia de gaps en decimal es: 24, 3, 7, 1, 11, 10, 9, 5

La correspondiente posting list es: [24, 27, 34, 35, 46, 56, 65, 70]

2)

Solución:

La secuencia de gaps en decimal es: [84, 8, 185, 299].

La correspondiente secuencia de bits utilizando codificación variable en bytes es:

11010100 10001000 00000001 10111001 00000010 10101011

3)

Solución:

I	W	D	E	I	I	O	E	F	W	O	F	E	I	O	C	O	D
O	F	E	I	O													
O		F	E	I	O												
		O	F	E	I	O											
			O	F	E	I	O										
				O	F	E	I	O									
					O	F	E	I	O								
										O	F	E	I	O			

4)

Solución:

Algunas diferencias entre HITS y pagerank:

- El método HITS calcula dos medidas HUB y AUTHORITY, el método PageRank solamente calcula una.
- HITS es capaz de encontrar buenas páginas independientemente de su contenido.
- A partir del resultado de una consulta HITS sólo hace análisis de enlaces, sin preocuparse del contenido.
- En HITS muchas páginas del resultado final pueden no contener ningún término de la consulta del usuario.
- En HITS, a partir de una consulta en un idioma se pueden recuperar páginas escritas en otros idiomas.
- El PageRank de cada página puede estar precalculado, HITS se debe calcular para cada consulta.

5)

Solución:

[0 0 1 0 0]

[0 1 0 0 0]

[0 1 0 1 1]

[0 1 0 0 0]

[0 0 0 0 1]

HUBS

[1.00 1.00 1.00 1.00 1.00]

[1.00 1.00 3.00 1.00 1.00]

[1.00 3.00 6.00 3.00 2.00]

[1.00 5.00 12.00 5.00 4.00]

[1.00 12.00 26.00 12.00 8.00]

AUTHS

[1.00 1.00 1.00 1.00 1.00]

[0.00 3.00 1.00 1.00 2.00]

[0.00 5.00 1.00 3.00 4.00]

[0.00 12.00 1.00 6.00 8.00]

[0.00 22.00 1.00 12.00 16.00]

[0.02 0.20 0.44 0.20 0.14]

[0.00 0.43 0.02 0.24 0.31]

