

# ACTO2 – SAR

(31/05/2021 - 2 puntos)

**(IMPORTANTE: Se deben justificar las respuestas, cálculos redondeados a 3 decimales)**

- 1) Se pide dar la secuencia de bits correspondientes a la compresión por códigos gamma de la siguiente posting list: [9, 16, 19, 20, 31, 41, 65, 71] **(0,3 puntos)**

Los gaps son: [9, 7, 3, 1, 11, 10, 24, 6]

Resultado final: 1110001 11011 101 0 1110011 1110010 111101000 11010

- 2) Esta pregunta consta de dos apartados: **(0,4 puntos)**

- a. Enuncia la ley de Heaps y justifica su utilidad en recuperación de información.

La ley de *Heaps* se formula de la siguiente forma  $M = kT^b$ , siendo K y b parámetros dependientes de la colección. Generalmente, en Recuperación de la Información, la ley de *Heaps* es usada para hacer una estimación del tamaño del vocabulario de la colección de documentos. Esta información es de gran relevancia en la implementación del índice invertido.

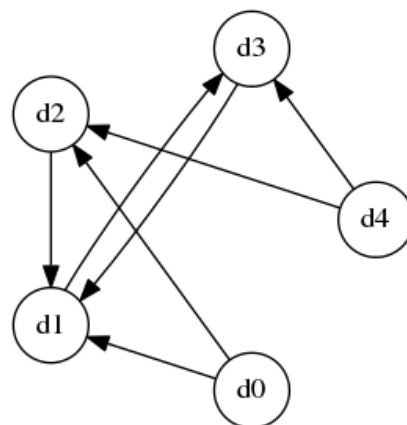
- b. Disponemos de dos colecciones diferentes de documentos que cumplen la ley de Heap. Suponiendo que ambas colecciones tienen el mismo tamaño en cuanto a número de tokens, ¿tendrán exactamente el mismo tamaño de vocabulario?

No. La ley de Heaps formula  $M = kT^b$ , siendo K y b parámetros dependientes de la colección. Por tanto, aunque las dos colecciones tengan el mismo tamaño para el número de tokens (T), el resultado puede dar distintos valores para M, dependiendo de los valores de los parámetros K y b de cada colección.

- 3) Dadas las siguientes páginas web y los enlaces entre ellas representadas como un grafo, se pide calcular el PageRank de cada página. Se debe calcular: i) la matriz de enlaces, ii) la matriz de probabilidades de transición, iii) la matriz de probabilidades de transición con teletransporte (utiliza un  $\alpha=0,1$  para el teletransporte), iv) todas las iteraciones para calcular el PageRank. Realiza como máximo cinco iteraciones, empezando como se indica:

$$\vec{x}_0 = (1.000, 0.000, 0.000, 0.000)$$

(0,8 puntos)



- i) la matriz de enlaces,

	D0	D1	D2	D3	D4
D0	0	1	1	0	0
D1	0	0	0	1	0
D2	0	1	0	0	0
D3	0	1	0	0	0
D4	0	0	1	1	0

- ii) la matriz de probabilidades de transición sin teletransporte

	D0	D1	D2	D3	D4
D0	0	0.5	0.5	0	0
D1	0	0	0	1	0
D2	0	1	0	0	0
D3	0	1	0	0	0
D4	0	0	0.5	0.5	0

- iii) la matriz de probabilidades de transición con teletransporte ( $\alpha=0,1$ )

	D0	D1	D2	D3	D4
D0	0.02	0.47	0.47	0.02	0.02
D1	0.02	0.02	0.02	0.92	0.02
D2	0.02	0.92	0.02	0.02	0.02
D3	0.02	0.92	0.02	0.02	0.02
D4	0.02	0.02	0.47	0.47	0.02

iv) las iteraciones para calcular el pagerank. Realiza como máximo 5 iteraciones.

	d0	d1	d2	d3	d4
t=0	1	0	0	0	0
T = 1	0.02	0.47	0.47	0.02	0.02
T = 2	0.02	0.47	0.038	0.452	0.02
T = 3	0.02	0.47	0.038	0.452	0.02

**SOLUCIÓN:**  $\pi = (0.02, 0.47, 0.038, 0.452, 0.02)$

- 4) Se pide indicar sobre la tabla, los desplazamientos que se realizarían en una búsqueda por Booyer-Moore del patrón “DOCOA” en la cadena “OOWAAAWIWBDOCOABWD”. Calcula el vector de desplazamiento correspondiente. **(0,5 puntos)**

O	O	W	A	A	A	W	I	W	B	D	O	C	O	A	B	W	D						
D	O	C	O	A																			
	D	O	C	O	A																		
		D	O	C	O	A																	
								D	O	C	O	A											
										D	O	C	O	A									

**Vector de desplazamiento:**

D	O	C	A	Otros
4	1	2	0	6