

ACTO2 – SAR

(21/05/2018 - 3 puntos)

(IMPORTANTE: Se deben justificar las respuestas, todos los cálculos se mostrarán redondeados a tres decimales)

- 1) Se pide indicar el mayor gap que puede ser codificado usando 1 byte: **(0,3 puntos)**

a) En un esquema de compresión por códigos gamma

Dado que la codificación en códigos gamma devuelve siempre una secuencia impar, de los 8 bits se utilizan realmente 7 bits. De los 7 bits, los tres primeros corresponderán a la longitud (111) el siguiente bit es 0 (el bit separador) y los tres siguientes para el mayor valor valdrán 111 y se corresponden con el offset. Por lo tanto el valor codificado es 1111, por lo que el mayor gap codificado con este esquema es un 15.

b) En un esquema de compresión por codificación variable en bytes

Con un único byte, el bit más significativo deberá estar puesto a 1, el resto, para codificar el mayor gap estará puesto a 1111111, por lo que el mayor gap codificado con este esquema es 127.

- 2) Se pide dar la secuencia de bits correspondientes a la compresión por codificación variable en bytes de la siguiente postings list: [513, 525, 22544] **(0,2 puntos)**

La secuencia de gaps es [513, 12, 22019] por lo que la codificación variable en bytes resulta:

00000100 10000001 10001100 00000001 00101100 10000011

- 3) En el cálculo del pagerank de una colección de páginas web: **(0,5 puntos)**

a. Justifica la introducción del teletransporte.

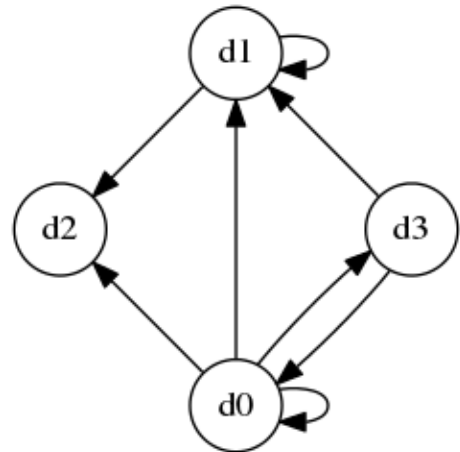
Asumimos que la web es un grafo dirigido que queremos modelar como una cadena de Markov y aplicar el cálculo del pagerank. Para que el pagerank de una página (su tasa de visitas a largo plazo) sea único, el modelo debe ser ergódico. El teletransporte se introduce para conseguir que el grafo de la web sea una cadena de Markov ergódica. El efecto del teletransporte es eliminar los ceros de la matriz inicial de probabilidades tanto en el caso de los callejones sin salida como en el resto de componentes.

b. ¿En qué intervalo toma valores el parámetro α , y cómo interesa que sea su valor?

El parámetro toma valores en el intervalo $[0,1]$. Interesa que su valor sea bajo para que el teletransporte no distorsione demasiado la distribución de probabilidad calculada en base a los enlaces, los valores habituales van entre 0,1 y 0,15.

- 4) 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 $\alpha=0,15$ para el teletransporte), iv) todas las iteraciones para calcular el pagerank. Realiza como máximo cinco iteraciones.

(1 punto)



Solución:

Matriz de enlaces:

```

[1 1 1 1]
[0 1 1 0]
[0 0 0 0]
[1 1 0 0]

```

Matriz de probabilidades de transición inicial:

```

[ 0.250  0.250  0.250  0.250]
[ 0.000  0.500  0.500  0.000]
[ 0.000  0.000  0.000  0.000]
[ 0.500  0.500  0.000  0.000]

```

Matriz de probabilidades de transición con teletransporte ($\alpha=0.15$):

```

[ 0.250  0.250  0.250  0.250]
[ 0.038  0.463  0.463  0.038]
[ 0.250  0.250  0.250  0.250]
[ 0.463  0.463  0.038  0.038]

```

Calculo del Pagerank:

```

x = [1 0 0 0]
x = [ 0.250  0.250  0.250  0.250]
x = [ 0.250  0.356  0.250  0.144]
x = [ 0.205  0.356  0.295  0.144]
x = [ 0.205  0.356  0.295  0.144]

```

Pagerank: $\pi = [0.205 \ 0.356 \ 0.295 \ 0.144]$

- 5) Se quiere buscar el patrón “CLOS” en el texto “BABICLONICOS”. Para ello se hace una búsqueda aproximada para obtener aquellos segmentos cuya distancia al patrón es menor o igual que 1. Se pide construir la matriz que corresponde al algoritmo de búsqueda aproximada e indicar las soluciones, es decir, los segmentos del texto que son resultados de la búsqueda. Se consideran pesos 1 para las Sustituciones, Inserciones y Borrados. **(1 punto)**

Solución

S	4	4	4	4	4	3	2	1	1	2	3	2	1
O	3	3	3	3	3	2	1	0	1	2	2	1	2
L	2	2	2	2	2	1	0	1	2	2	1	1	2
C	1	1	1	1	1	0	1	1	1	1	0	1	1
	0	0	0	0	0	0	0	0	0	0	0	0	0
		B	A	B	I	C	L	O	N	I	C	O	S

Subcadenas:

CLO (5-7)
CLON (5-8)
COS (10-12)