

(31/05/2021 - 2 puntos)

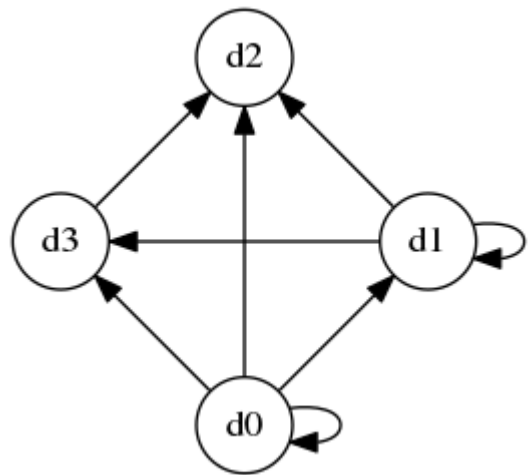
(IMPORTANTE: Se deben justificar las respuestas, cálculos redondeados a 3 decimales)

- 1) Se pide dar la secuencia de bits correspondientes a la compresión por códigos gamma de la siguiente posting list: [7, 12, 13, 16, 20, 26, 34, 35, 43]: **(0,3 puntos)**
- 2) Esta pregunta consta de dos apartados: **(0,4 puntos)**
 - a. Enuncia la ley de Heap y justifica su utilidad en recuperación de información.
 - 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?

- 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:

$$\overrightarrow{x_0} = (1.000, 0.000, 0.000, 0.000)$$

(0,8 puntos)



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

[illegible]

SOLUCIÓN:

1)

La secuencia de gaps en decimal es: [7, 5, 1, 3, 4, 6, 8, 1, 8]

La correspondiente secuencia de bits es:

11011 11001 0 101 11000 11010 1110000 0 1110000

2)

a. La ley de Heap dice $M = kT^b$, donde K y b son parámetros que dependen de la colección, T es el número de tokens y M el tamaño del vocabulario de la colección. En RI se utiliza para hacer una estimación del tamaño del vocabulario de la colección de documentos, M . Esta información es importante cuando se aborda la implementación del índice invertido.

b. No

La ley de Heap dice $M = kT^b$, donde K y b son parámetros que dependen de la colección. Por tanto, aunque las dos colecciones tengan el mismo valor para el parámetro T , el resultado de aplicar la ley puede proporcionar valores diferentes para M , dependiendo de los valores de los parámetros K y b de cada colección.

3)

Matriz de enlaces

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

Matriz de probabilidades de transición

```
[[ 0.250 0.250 0.250 0.250]
 [ 0.000 0.333 0.333 0.333]
 [ 0.000 0.000 0.000 0.000]
 [ 0.000 0.000 1.000 0.000]]
```

Matriz de probabilidades de transición con teletransporte

```
[[ 0.250 0.250 0.250 0.250]
 [ 0.025 0.325 0.325 0.325]
 [ 0.250 0.250 0.250 0.250]
 [ 0.025 0.025 0.925 0.025]]
```

Iteraciones cálculo de PageRank

```
 $\vec{x}_0 = (1.000, 0.000, 0.000, 0.000)$ 
 $\vec{x}_1 = (0.250, 0.250, 0.250, 0.250)$ 
 $\vec{x}_2 = (0.138, 0.213, 0.438, 0.213)$ 
 $\vec{x}_3 = (0.155, 0.219, 0.410, 0.219)$ 
 $\vec{x}_4 = (0.152, 0.218, 0.415, 0.218)$ 
 $\vec{x}_5 = (0.153, 0.218, 0.414, 0.218)$ 
 $\vec{\pi} = (0.153, 0.218, 0.414, 0.218)$ 
```

4)

Vector de desplazamiento

I	E	O	C	F	A	W	D
0	5	5	1	2	5	5	3

[illegible]