

ACTO2 – SAR

(19/06/2017 - 3 puntos)

(IMPORTANTE: todos los cálculos se mostrarán redondeados a dos decimales)

- 1) Se pide obtener la postings list a partir de la siguiente secuencia de bits codificada utilizando codificación variable en bytes: **(0,5 puntos)**

00000010 10000100 10001011 00000001 00010100 10001101

Solución:

Los valores codificados son $[(2*128)+4=260, 11, [(1*128*128)+(20*128)+13=18.957]]$.

Recordando que los valores codificados corresponden a los gaps, la postings list es [260, 271, 19228]

- 2) Asumiendo que la longitud de la postings list de una colección de documentos sigue una ley de Zipf $\sim i^{-2}$, y que las primeras 100 listas más largas tienen una longitud ≥ 1000 , ¿qué longitud tiene la lista más larga? **(0,5 puntos)**

Solución:

La ley de Zipf: $c_{fi} = K/i^2$

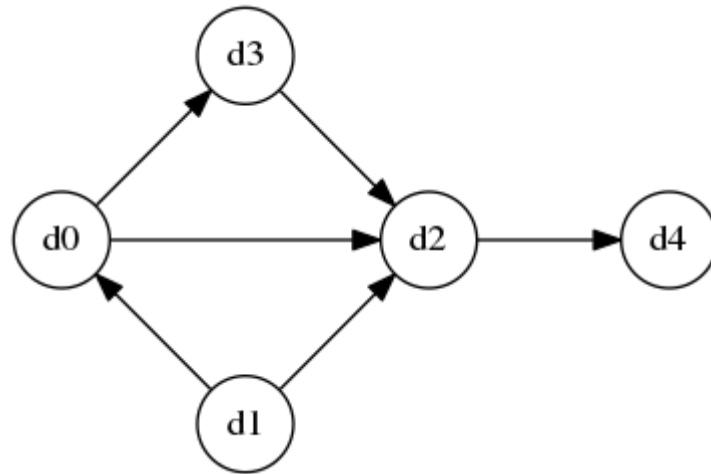
Para $i = 100$, $cf_i = 1.000$ por lo que $K = i^2 * cf_i = 100^2 * 1.000 = 10^7$.

Para $i = 1$, $cf_i = K/i^2$, es decir, $cf_i = K$, por lo que $cf_1 = 10^7$, que es la longitud de la lista más larga.

- 3) Se pide indicar sobre la tabla, los desplazamientos que se realizarían en una búsqueda por Booyer-Moore del patrón "FCFAAB" en la cadena "CFFFCABBBEBFCFAABEBAF". **(0,75 puntos)**

[illegible]

- 4) 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 cinco iteraciones sin normalización. **(1,25 puntos)**



Solución:

Matriz de enlaces:

```

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

HUBS

AUTHORITY

t0	[1 1 1 1 1]	[1 1 1 1 1]
t1	[2 2 1 1 0]	[1 0 3 1 1]
t2	[4 4 1 3 0]	[2 0 5 2 1]
t3	[7 7 1 5 0]	[4 0 11 4 1]
t4	[15 15 1 11 0]	[7 0 19 7 1]
t5	[26 26 1 19 0]	[15 0 41 15 1]