

Acto2 - SAR

(16/05/2016 - 3 puntos)

Apellidos y Nombre:.....

(IMPORTANTE: todos los cálculos se mostrarán truncados 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)**

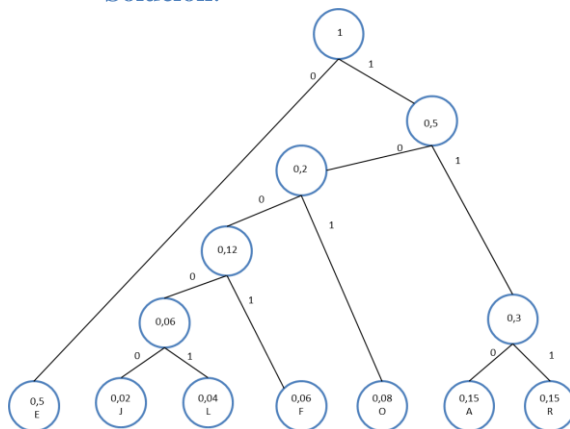
00000110 10010100 10001010 00000001 00010100 10001101

Solución:

Los valores codificados son $[(6*128)+20=788, 10, [(1*128*128)+(20*128)+13=18.957]$. Recordando que los valores codificados corresponden a los gaps, la postings list es [788, 798, 19.755]

2) Dada la siguiente lista de símbolos y sus probabilidades se pide construir el árbol de Huffman: **(0,5 puntos)**

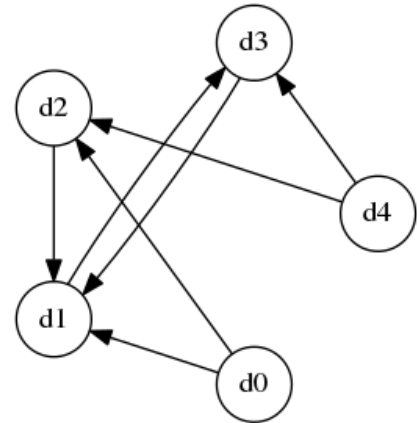
Solución:



Símbolo	Probabilidad
A	0.15
E	0.50
F	0.06
J	0.02
L	0.04
O	0.08
R	0.15

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.

(1 punto)



Solución:

Matriz de enlaces:

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

Matriz de probabilidades de transición sin teletransporte:

```
[0.000 0.500 0.500 0.000 0.000]
[0.000 0.000 0.000 1.000 0.000]
[0.000 1.000 0.000 0.000 0.000]
[0.000 1.000 0.000 0.000 0.000]
[0.000 0.000 0.500 0.500 0.000]
```

Matriz de probabilidades de transición con teletransporte ($\alpha=0.1$):

```
[0.020 0.470 0.470 0.020 0.020]
[0.020 0.020 0.020 0.920 0.020]
[0.020 0.920 0.020 0.020 0.020]
[0.020 0.920 0.020 0.020 0.020]
[0.020 0.020 0.470 0.470 0.020]
```

Cálculo del Pagerank

Redondeando a 3 decimales:

```
 $\vec{x}_0 = (1.000, 0.000, 0.000, 0.000, 0.000)$ 
 $\vec{x}_1 = (0.020, 0.470, 0.470, 0.020, 0.020)$ 
 $\vec{x}_2 = (0.020, 0.470, 0.038, 0.452, 0.020)$ 
 $\vec{x}_3 = (0.020, 0.470, 0.038, 0.452, 0.020)$ 
```

$\vec{\pi} = (0.020, 0.470, 0.038, 0.452, 0.020)$

Truncando con 2 decimales:

```
 $\vec{x}_0 = (1.00, 0.00, 0.00, 0.00, 0.00)$ 
 $\vec{x}_1 = (0.02, 0.47, 0.47, 0.02, 0.02)$ 
 $\vec{x}_2 = (0.02, 0.47, 0.03, 0.45, 0.02)$ 
 $\vec{x}_3 = (0.01, 0.46, 0.03, 0.45, 0.01)$ 
 $\vec{x}_4 = (0.01, 0.45, 0.02, 0.43, 0.01)$ 
 $\vec{x}_5 = (0.01, 0.42, 0.02, 0.42, 0.01)$ 
```

4) Se quiere buscar el patrón “SODA” en el texto “SSOLASLASEDAS”. Para ello se hace una búsqueda aproximada para obtener aquellos segmentos cuya distancia (de Levenshtein) al patrón es menor o igual que 1. Se pide completar la siguiente matriz que corresponde al algoritmo de búsqueda aproximada e indicar las soluciones, es decir, los segmentos del texto que son resultados de la búsqueda. **(1 punto)**

Para el cálculo de la distancia se consideran pesos 1 para las Sustituciones, Inserciones y Borrados.

Solución:

A	4	3	3	2	2	1	2	3	2	3	3	3	2	1	2			
D	3	2	2	1	1	2	2	2	2	2	2	2	1	2	2			
O	2	1	1	0	1	2	1	1	2	1	1	1	2	2	1			
S	1	0	0	1	1	1	0	1	1	0	0	1	1	1	0			
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
		S	S	O	L	A	S	L	A	S	S	E	D	A	S			

Segmentos encontrados con distancia 1:

“SOLA” de la posición 2 a la 5.

“SEDA” de la posición 10 a la 13.