

## Organización de Datos – Curso Servetto

Evaluación Módulo Recuperación de Textos, 26 de Noviembre de 2007

Documentos:

Nº Documento	Texto Documento
1	La vida es negro, yeh
12	El rock es negro, uoh
15	La nena es negro, yeah!
40	El rock es nena uoh
60	El rock y el negro son nena
100	Yeah, aye, aye, Uoh, no

(Considerando que cada palabra mayor a 3 caracteres un término y cada palabra un documento)

- 1) Suponiendo que se deben comprimir los punteros. Realice un análisis para reconocer la mejor codificación posible entre Delta, Gamma y Unario para este caso en particular.
- 2) Realice un índice invertido por Sort de los términos.
- 3) En el índice del punto anterior realice una mejora concatenando términos. Indique cómo queda el índice.
- 4) Genere un índice del tipo Signature Files y, sobre él, realice la consulta “soy rock, nena”. Y pondere el resultado (ranking) con un método de Producto Interno.

## Organización de Datos – Curso Servetto

Evaluación Módulo Compresión, 26 de Noviembre de 2007

- 1) Si tengo que comprimir los mensajes transmitidos por una fuente que se limita a “decir”:  
“Nacho Pancho, Nacho Pancho”  
Límites: Sólo se cuenta con un Lz77, con ventanas de trabajo y observación de 12 lugares cada una, y un Aritmético de Orden 0.  
Analice cuál es el que alcanza el mayor nivel de compresión.
- 2) Comprima con un PPMC de Orden 2 el siguiente Texto:  
“oyeahayeayeuoh” (Obs: No olvide ninguna parte del PPMC)
- 3) Determine si el siguiente archivo puede haber sido generado con un algoritmo de Block Sorting, Move To Front y Shannon. De ser posible, muestre el archivo original.  
“1001000050000410101001000100006”  
Suponiendo que el 257 es el código del EOF y que el espectro de mensajes posibles está acotado a “abcdeEOF”.
- 4) Explique los pasos para realizar una compresión de Huffman con el manejo eficiente del árbol. Dé un ejemplo.