

Organización de Datos – Curso Servetto – Evaluación de Recuperación de Textos – 19/8/09

1. Codificar la siguiente lista invertida con el método de Golomb, sabiendo que el sistema cuenta con 1318 documentos:
312, 372, 417, 609, 676, 1315
2. Dada la siguiente secuencia de términos del léxico de un sistema de recuperación de textos, mostrar cómo quedaría codificada en un nodo de un árbol B+ usando abreviaturas frontales parciales (el símbolo # representa la dirección de almacenamiento de la lista invertida): (moises, 10, #), (mojada, 1, #), (mojado, 2, #), (moje, 1, #), (molares, 1, #), (moldavia, 1, #), (molde, 1, #), (moldeado, 1, #), (moldes, 2, #), (molecula, 1, #), (moledora, 1, #), (molesta, 9, #), (molestaba, 3, #), (molestado, 1, #), (molestan, 4, #), (molestar, 2, #), (molestara, 1, #), (molestaron, 1, #), (molestas, 3, #), (molestes, 1, #), (molestia, 60, #), (molestias, 9, #), (molesto, 19, #), (molestos, 7, #)

Analice cómo se podría reducir la lista explicando las técnicas que proponga y mostrando cómo quedaría la lista. Analice también la inconveniencia aplicarlas buscando un ejemplo de consulta.

3. Describa las métricas que conoce para evaluar un sistema de recuperación de textos y explique cómo las evaluaría.