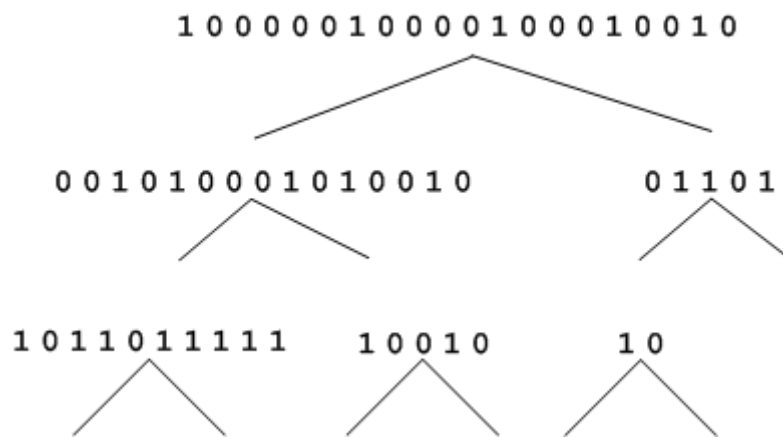


O seguinte boletín de exercicios deberase enviar ao profesor Antonio Fariña á dirección (antonio.lbd@gmail.com) antes do **día 17 de abril de 2012**.

11. Indíquense as principais vantaxas dos *Dynamic Lightweighth Dense Codes* fronte aos seus análogos que non son *lightweight*.
12. Por que é de esperar que un código semistático sexa máis rápido ao descomprimir que un dinámico. Análogamente, indíquese por que os códigos semistáticos son os máis axeitados cando se require a posibilidade de facer buscas directas sobre o texto comprimido. ¿Sería ilóxico pensar que unha técnica dinámica consiga obter os mesmos tempos de busca que unha técnica semistática? Razónense as respostas.
13. Cales son as principais diferencias entre unha técnica de compresión “*fixed-to-variable*”, “*variable-to-fixed*”, e “*variable-to-variable*”? Indica algún representante para cada un dos tres tipos de técnicas. Indica brevemente en que se basean ditas alternativas para conseguir compresión. ¿Coñeces algunha técnica “*fixed-to-fixed*”?
14. Indica a secuencia final comprimida que se obtería ao aplicar BPE sobre a secuencia: GABCDABABBAFAFABCDABCDEF. Supóñase que os seguintes caracteres libres (a utilizar nas substitucións do BPE) son: {H,I,J,K,L,M,N...}
15. Calcula a transformación Burrows-Wheeler (T^{BWT}) para o texto seguinte: T=“ba_a_labar_a_la_alabarda\$” (nota: o carácter ‘_’ indica un espazo en branco).
16. Explica brevemente as principais diferencias entre un índice tradicional e un autoíndice, ten en conta as operacións de count/locate/extract na túa explicación. ¿Deben ser capaces de soportar as mesmas operacións á hora de realizar buscas? ¿Que operacións deben soportar e cal é a súa función?
17. Tendo en conta a estrutura básica dun índice invertido que indexa un texto dado:
 - vocabulario e listas invertidas para cada termo, por unha banda
 - texto comprimido ou sen comprimir, por outra...
 - (a) Indíquense os pasos que son necesarios para saber cantas veces aparece a frase “white house” no texto indexado. *Diferenciese* o caso dun índice invertido tradicional (con información posicional completa) fronte a un índice invertido con direccionamento por bloques.
 - (b) Explíquese brevemente se o rendemento dun índice invertido con direccionamento por bloques que indexe texto comprimido (con Tagged Huffman, ETDC ou SCDC) debería superior ao dun índice invertido análogo que indexe texto sen comprimir á hora de localizar ocorrencias dun patrón dado. ¿Por que?
 - (c) Se temos dous índices invertidos X e Y con direccionamento por bloques, usando X bloques de 1024 bytes e Y bloques de 16384 bytes, e estando o texto indexado comprimido con Tagged Huffman... Indica se as seguintes afirmacións son verdadeiras ou falsas (razona a túa resposta).
 - X ocupa máis que Y.
 - X é capaz de resolver máis rapidamente que Y consultas que afecten a palabras que ocorren unha soa vez. ¿Por que?
 - Se xa temos localizadas todas as aparicións dun patrón P, e queremos descomprimir un *snippet* de texto (50 palabras antes e 50 despois de cada ocorrencia) para cada unha das ocorrencias... ¿En que caso se realizará antes dita operación?
18. Se nun índice invertido comprimimos cada lista invertida (para cada termo) cunha técnica que non permite acceso directo:

- (a) ¿qué tipo de algoritmo de intersección de listas deberíamos utilizar?
- (b) Se asumimos que a operación máis frecuente que vai ter que facer o noso índice á a de buscar frases compostas de 2 palabras, onde a primeira palabra “A” aparece 1000 veces máis que a segunda palabra “B” (isto é, $frec_A = 2000 * frec_B$)...
 - a. ¿Segue a ser axeitado o uso do algoritmo de intersección anterior?
 - b. Se pensásemos en usar un algoritmo tipo *svs* con busca exponencial sobre a lista máis longa, como deberíamos modificar a representación das listas invertidas comprimidas para facelo posible?

19. Dado o alfabeto $\Sigma = \{_, a, b, c, d, l, r\}$ e o Wavelet-tree seguinte, indíquese cal foi o texto orixinal que está sendo “autoindexado” (é dicir, o texto sobre o que se creou o wavelet-tree)



20. Dado o texto “va a lavar o a alabar a la alabarda\$” (onde \$ indica un caracter alfanumérico menor que calquera outro caracter do texto)
- (a) constrúase o Array de sufixos correspondente a dito texto.
 - (b) Constrúanse os vectores Ψ e D necesarios para obter un autoíndice comprimido CSA (Sadakane).