

Examen Parcial de Recuparació de la Informació
Curs 2006-2007, Quadrimestre d'hivern

Temps: 2h

Exercici 1 (2 punts)

- Simuleu la cerca del patró *RTAGG* dins la seqüència *AGRATARGATTC*, on *R* és el caràcter extés que vol dir *A* o *G*, aplicant l'algorisme BNDM.

- Quin seria el cost teòric espacial del primer apartat.

Exercici 2 (2 punts)

- Simuleu la cerca del patró *CCBCBA* dins la seqüència *ABCACBACCBCBAT*, aplicant l'algorisme BOM.

- Existeix algun mot acceptat per l'autòmata però que no sigui factor?

Exercici 3 (1 punt) Apliqueu l'algorisme de programació dinàmica, amb es valors típics per defecte, per trobar el millor aliniament entre els dos clusters següents: el primer de dues seqüències (a) i (ac) alineades ja com

a -
a c

i el segon de dues seqüències (a) i (c) alineades

a
c

Exercici 4 (1 punt)

Donades les similaritats entre les sis seqüències S_1, \dots, S_6 troba l'arbre de l'aliniament múltiple.

	S_1	S_2	S_3	S_4	S_5	S_6
S_1		20	14	16	-10	12
S_2			12	14	14	6
S_3				18	16	16
S_4					16	2
S_5						6

Exercici 5 (2 punts) Un jugador juga contra el casino segons el resultat de tirar un dau (de 4 cares: un tetraedre) i guanya qui ha tret un resultat més alt. El jugador sempre juga amb un dau equilibrat pero el casino en té dos, un d'equilibrat i un altre on el 4 té la probabilitat de $1/2$ i la resta de valors 1,2,3 tenen la mateixa, i en cada tirada el casino escull un del dos daus a l'atzar excepte en la primera que agafa el dau equilibrat.

- Quina és la probabilitat de que el casino tregui els resultats 4,4,4?

- Quants cops més gran és aquesta probabilitat que si hagués jugat tota l'estona amb un dau equilibrat?

Exercici 6 (0.5 punts per apartat)

- – Donat un text aleatori sobre l'alfabet $\{a, c, g, t\}$, quin és el valor esperat del salt de la finestra si el patró que estem buscant és *tcaa* i l'algorisme el de Horspool.

– I si la distribució del text fora: a:40%, c:20%, g:20% i t:40%?

- Donat un text de longitud n , quin és el cost en espai i temps de construir el suffix array? I el cost temporal de buscar un patró de longitud l ?

- Explica en quins casos és aplicable el mètode de seqüenciació per hibridació i perquè.

- Si haguessis de buscar un milió de patrons de DNA de longitud entre 10 i 15 bases,
 - quin algorisme utilitzaries i perquè.

- Podries implementar-lo en un PC amb 2Gb de RAM?