

Examen Parcial de Recuparació de la Informació
Curs 2007-2008, Quadrimestre d'hivern

Temps: 2h

Exercici 1 (1.5 punts)

- Simuleu la cerca del patró *CGCT* a la seqüència *CGATATGGGCT* aplicant l'algorisme de Horspool.

- Quin seria el número esperat de salts al buscar aquest patró en una seqüència aleatòria equiprobable de 11000 caràcters?

Exercici 2 (1.5 punts)

- Simuleu la cerca dels patrons *ATAA*, *AATA*, *TATATA*, *TATAA* dins la seqüència *AATTATATAA*, aplicant l'algorisme SBOM.

Exercici 3 (1.5 punts)

- Simuleu la cerca del patró *RARG* dins la seqüència *AARTARRA* on *R* és un caràcter extés, aplicant l'algorisme BNDM.

- Si aplicant l'algorisme BNDM per al mateix patró *RARG* obtenim $D_1 = (1011)$, $D_2 = (0110)$, $D_3 = (1100)$ i $D_4 = (1111)$, quants textos són possibles?

Exercici 4 (1.5 punts)

- Trobeu el millor alineament entre les 3 seqüències *AAT*, *ACA* i *AT*.

Exercici 5 (1.5 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, que és amb el que comença a jugar, i un desequilibrat on cada valor i ($i = 1, 2, 3, 4$) té la probabilitat de $i/10$, i el casino canvia de dau en promig cada cinc tirades.

- Quina és la seqüència d'estats més probable si els resultats han estat 1434.

- Quina probabilitat té el casino de treure un 4?

Exercici 6 (1.5 punts)

- Si un investigador us diu que de tant en tant ha de buscar patrons en el genoma humà (3Gb) què li proposariau?

- i si diu que diàriament li cal buscar mil.lers, què li proposariau?

- Un investigador us demana un algorisme per trobar el millor aliniament entre seqüències de longitud 10, i us afegeix que a vegades en té 5, altres 10 i poc sovint més de deu que poden arribar a 30. Què li diriau?

- Un investigador vol saber la freqüència de tots els patrons de longitud 8 en el genoma

humà. Com us ho fariau?