

**Examen Parcial de Recuparació de la Informació**  
**Curs 2007-2008, Quadrimestre de primavera**

**18 de juny de 2009. Temps: 2h**

**Exercici 1 (1.5 punts)**

- Simuleu la cerca del patró *RARG* dins la seqüència *AARCAGRGGTT* on  $R = \{A, G\}$  és un caràcter extés, aplicant l'algorisme BNDM.

- Si aplicant l'algorisme BNDM per al patró *CGRGA* obtenim  $D_1 = (11111)$ ,  $D_2 = (01010)$ ,  $D_3 = (10100)$  i  $D_4 = (0000)$ , quins textos sobre l'alfabet extés  $\{A, C, G, T, R\}$  són possibles?

**Exercici 2** (1.5 punts)

- Simuleu la cerca del patró  $RYRY$  dins la seqüència  $ACCAAYGC$ , on  $R = \{A, G\}$  i  $Y = \{C, T\}$ , aplicant l'algorisme de Horspool.

- Si busquessim els patrons  $ACCT$ ,  $ACTA$  i  $CGAT$  amb una taula de dos símbols, quin seria el salt esperat sobre un text aleatori amb una distribució equiprobable?

**Exercici 3** (2 punts) Simuleu la cerca dels patrons  $AAT$ ,  $TTTA$ ,  $TTCTA$  dins la seqüència  $ATCCTTAT$  aplicant l'algorisme SBOM (dibuixeu l'autòmata amb totes les transicions i els suffix-links).

**Exercici 4** (1.5 punts)

Apliqueu l'algorisme de programació dinàmica per buscar totes les aparicions del patró  $ATT$  amb un error dins la seqüència  $CATTC$ .

**Exercici 5** (2 punts)

- Si un investigador us diu que de tant en tant ha de buscar patrons en el genoma humà (3Gb), quin algorisme li proposarieu i per què? Li podeu donar una estimació del temps de cada cerca?

- i si diu que diàriament li cal buscar mil.lers, quin algorisme li proposarieu i per què? Serien útils els suffix trees? i els suffix arrays?

- 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 i d'altres 10. Què li dirieu?

- Un investigador vol saber la freqüència de tots els patrons de longitud 8 en el genoma humà. Com us ho farieu?

**Exercici 6** (1.5 punts) Un jugador juga contra el casino segons el resultat de tirar un dau de 4 cares (un tetraedre) i perd qui ha tret un resultat més baix. El jugador sempre juga amb un dau equilibrat pero el casino en té tres, un d'equilibrat i dos altres on cada valor  $i$  ( $i = 1, 2, 3, 4$ ) té la probabilitat de  $i/10$ , i en cada tirada el casino escull un dels tres daus a l'atzar.

- Quina és la seqüència d'estats més probable si els resultats del casino han estat 33.

- En cada tirada quina és la probabilitat de que perdi el casino?