

**Examen Parcial de Recuparació de la Informació**  
**Curs 2009-2010, Quadrimestre de primavera**

**Temps: 2h**

**Exercici 1** (1.5 punts)

- Simuleu la cerca del patró *AAT* dins la seqüència *ATCTAATTCC*, aplicant l'algorisme BNDM.

- Calcula el valor esperat del salt.

**Exercici 2** (1.5 punts) Simuleu la cerca dels patrons  $AAT$ ,  $ATA$ ,  $TTAT$  dins la seqüència  $ATCTAATTAT$  aplicant l'algorisme SBOM (dibuixeu l'autòmata amb totes les transicions i els suffix-links).

**Exercici 3** (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 4** (1.5 punt) Apliqueu l'autòmata de Thompson per buscar l'expressió regular  $A(A^*|G^*)G$  dins la seqüència  $GAGGA$ .

**Exercici 5** (2 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 dos, un d'equilibrat i un altre on cada valor  $i$  ( $i = 1, 2, 3, 4$ ) té la probabilitat de  $i/10$ , i en cada tirada el casino escull un del dos daus a l'atzar.

- Quina és la seqüència d'estats més probable si els resultats del casino han estat 14.

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

**Exercici 5** (0.5 punts cada apartat)

- Has creat una seqüència de valors amb el model

$M = 6$

$N = 2$

A:

0.5 0.5

0.5 0.5

B:

0.5 0.1 0.1 0.1 0.1 0.1

0.1 0.1 0.1 0.1 0.1 0.5

pi:

0.5 0.5

Quin creus que seria el model que *esthm* t'estimaria i perquè?

- Si hem de buscar molts patrons en una seqüència, quins són els paràmtres a tenir en compte per escollir el millor algorisme?

- Perquè no és possible buscar el millor aliniament entre més de 10 seqüències llargues?

- Perquè s'utilitzen suffix-trees per buscar MUMs?