Contrôle continu & Examen final LI014
(Traitement Automatique de l'Écrit)
Session de juin
Durée : 3 heures.
Aucun document autorisé.

1. **Généralités.** Comparez brièvement les avantages et inconvénients des approches symboliques et non symboliques en TAL.
2. **Cours de Leo Wanner.** Soient les quatre documents suivants :
   (a) Le gîte de La Haute Mancelière est très joli et a été récemment rénové (jusqu'à 7 personnes + bébé). Le gîte est classé 3 épis auprès des Gîtes de France. Belle campagne, plages de sable, nombreux sîtes historiques (Mont St Michel, St Malo, Rennes, Fougères, Vitré, Combourg...). Celui-ci est ouvert toute l'année.
   (b) À proximité de la mer (5 à 10 km) dans un cadre champêtre ces gîtes sont une excellente escale pour explorer la côte de granit rose au rythme des marées. De nombreux sentiers pédestres sillonnent la côte des ajoncs ou la presqu'île sauvage dans une nature encore préservée avec un riche patrimoine culturel et artistique à découvrir.
   (c) En Bretagne, maison de campagne sur la côte d'émeraude à 3 km de la mer, proche de Dinard (10 kms), Saint-Malo et Dinan (15 kms). Idéal pour tous ceux qui recherchent confort, calme et espace à proximité des plages et des sites touristiques. Location toute l'année à la semaine ou au week-end.
   (d) A 200m de la mer, dans le Finistère, deux gîtes ruraux tout confort situés sur un ancien corps de ferme. Capacité : 4/5 personnes pour l'un et 7 personnes pour le second gîte. Environnement calme. Gîtes 3 épis.

   1.1 Quelles opérations textuelles connaissez-vous ? Appliquez ces opérations au document (2a).
   1.2 Construire une structure d'index appropriée pour le document (2b).
   1.3 Supposons que vous souhaitiez trouver un document qui contient de l'information à propos de X ou Y mais pas de Z, ou bien ne contient aucune information à propos de A. Formulez la requête booléenne correspondante.
   1.4 Transformez la requête de la question 1.3 en forme normale disjonctive (et construisez-en la table de vérité).
   1.5 Soit l'espace vectoriel des pondération de termes suivant :

	d_1	d_2	d_3	d_4
nature	0	1	0	0
mer	0	1	1	1
campagne	1	0	1	0
plage	1	0	1	0
St Malo	1	0	1	0

   Soit de plus la requête *nature* $\wedge$ *mer* $\wedge$ *campagne* $\wedge$ *plage* $\wedge$ St Malo avec les pondérations suivantes :

nature	mer	campagne	plage	St Malo
0.5	2	1	2	3

   Supposez que ne soient fournis à l'utilisateur que les documents de poids total supérieur à $4,0$. Quels documents l'utilisateur recevra-t-il, et dans quel ordre ?
3. **Langages formels**
   (a) Soit la grammaire formelle $\mathcal{G}_1 = \begin{array}{l} S \longrightarrow aSa \mid bTb \mid U \\ T \longrightarrow cSc \\ U \longrightarrow z \end{array}$

   Pour un mot engendré que vous choisirez, donnez (1) un exemple de chaîne de dérivation et (2) l'arbre syntaxique correspondant. Soit $L_{\mathcal{G}_1}$ le langage engendré par $\mathcal{G}$.
   (b) Soit l'expression rationnelle $a^*b^*c^*zc^*b^*a^*$. Soit $L_{\mathcal{E}}$ le langage correspondant à cette expression. Lequel des langages $L_{\mathcal{G}_1}$ et $L_{\mathcal{E}}$ est-il inclus dans l'autre ?
   (c) Proposez un automate qui reconnaît $L_{\mathcal{E}}$.