Cours de Bases de données
Examen écrit - Juin 2007

Livres fermés. Durée : 3 heures 1/2.
Répondez à chaque question sur une feuille séparée (hormis le QCM)
sur laquelle figure vos nom, prénom, et section.

Soyez clairs, concis, et précis.

Remarque : Les étudiants en géomatique ne doivent pas répondre à la question 5.

1. *Modélisation*
   Le système de gestion d'un hôpital reprend les informations suivantes concernant les médecins, les patients et leur historique de traitement.
   - Chaque médecin est identifié par un numéro unique, attribué par l'hôpital. On connait également ses nom, prénom et adresse (qui est composée de la rue, du numéro dans la rue, du code postal et de la localité).
   - Les médecins sont attachés à un ou plusieurs départements. Chaque département est identifié par un numéro unique, et possède un libellé. Un département n'existe que si au moins un médecin y travaille. Un médecin est également désigné comme étant le chef du département. Il peut, le cas échéant, diriger plusieurs départements.
   - Les patients sont identifiés par un numéro unique, attribué par l'hôpital. On connait également leur nom, prénom, adresse, ainsi que le numéro de leur carte SIS. Il est possible qu'un médecin soit aussi un patient.
   - L'hôpital utilise un système de codification des maladies, un code, le nom scientifique et l'appellation commune étant associé à chaque maladie répertoriée. Un code par défaut est utilisé pour les cas non encore diagnostiqués.
   - On désire conserver un historique des soins effectuées au sein de l'hôpital. Ainsi, pour chaque hospitalisation, on enregistre la date de début et de fin, l'identifiant du patient et du médecin le traitant[1], ainsi que de la (des) maladie(s) pour lesquelles le patient a été soigné.

   (a) Dessinez un diagramme entités-relations conforme à la description ci-dessus. Précisez les clés des ensembles d'entités et des relations, ainsi que les contraintes d'intégrité éventuellement non représentées dans le diagramme ER.

   (b) Effectuez la conversion de ce diagramme vers le modèle relationnel.

2. *Théorie des dépendances*
   Soit *R(A,B,C,D,E)*, un schéma de relation et $F = \{AB \rightarrow C,\ CE \rightarrow D,\ CD \rightarrow B,\ C \rightarrow E\}$ un ensemble de dépendances fonctionnelles associées à R.

   (a) La décomposition *e(ABCD,ADE)* est-elle sans perte par rapport à F ? Conserve-t-elle les dépendances ? Justifiez.

   (b) Si la composition citée ci-dessus n'est pas sans perte, appliquez l'algorithme de décomposition en BCNF vu au cours. La nouvelle décomposition est-elle sans perte ? Justifiez.

[1] Si un médecin est hospitalisé, in ne peut évidemment se soigner lui-même.

3. *Langages d'interrogation*
   Une base de donnés de bioinformatique contient (entre autres) les relations suivantes :

   - *Gene(Id_gene, Start, End)* fournit le numéro d'identification des gènes ainsi que leur position sur le génome.
   - *Transcrit(Id_transcrit, TrStart, TrEnd, Id_gene)* contenant les informations relatives aux transcrits.
   - *Proteine(Id_proteine, PrStart, PrEnd, Id_transcrit)* contenant les informations relatives aux protéines.
   - *Orthologue(Id_gene, Id_gene_other)* contenant la liste des gènes correpondants dans les autres espèces.

   Soient les trois requêtes suivantes.

   (a) Donner le gène correspondant à la protéine "ENSP00000260950".

   (b) Donner, pour chaque protéine, la différence de taille entre le gène correspondant, et chacun des orthologues de ce gène (en indiquant la référence des gènes pour chaque différence).

   (c) Sachant qu'une 3'UTR est la partie située entre la fin d'une protéine et la fin du transcrit correspondant, donnez la moyenne de la taille des 3'UTR

   Exprimez les les trois requêtes en SQL.

4. *Implémentation du modèle relationnel*

   (a) Quel est l'intérêt de la technique des B-trees ? Que gagne-t-on en l'utilisant ?

   (b) Quand est-il intéressant d'utiliser une table "hash" pour calculer un joint ?

   (c) Donnez les principes généraux de l'optimisation des requêtes.

5. ** Bases de données déductives*

   (a) De quoi est composée une base de données déductive ?

   (b) Comment calcule-t-on l'extension d'une rêgle récursive ? Peut-on toujours exprimer une règle récursive en algèbre relationnelle ? Justifiez.

   (c) Sachant que, dans une BDD, il existe un prédicat *arc(S1,S2)* qui indique qu'il existe un arc entre S1 et S2, donner le prédicat *boucle* qui indique qu'il existe une suite d'arcs non vide entre un noeud et lui-même.

6. *Travaux pratiques*

   (a) Expliquez la différence entre PHP et HTML. Que permet l'un que ne permet pas l'autre ?

   (b) Décrivez en quelque mots le sujet du travail réalisé.

---

* Les étudiants en géomatique ne doivent pas répondre à cette question.