1. You are given two relation schemas, EMPLOYEE and DEPARTMENT. EMPLOYEE has the attributes ENUM (employee number), ENAME (employee name), DNUM (department number of the department in which the employee works in) and SALARY (employee salary in pounds). DEPARTMENT has the attributes DNUM (department number), DNAME (department name) and MGR (employee number of department manager).

   Answer the following three parts.

   a. Define the concepts of foreign key and referential integrity using the EMPLOYEE and DEPARTMENT schemas to illustrate your definitions.

   [11 marks]

   b. Define the concept of an ISA relationship used in an Entity-Relationship Diagram (ERD) using the EMPLOYEE and DEPARTMENT schemas to illustrate your definition.

   What is the connection between ISA relationships and referential integrity?

   [11 marks]

   c. An **inclusion dependency** (IND) is a statement of the form R[X] $\subseteq$ S[Y], where R and S are relation schemas and X and Y are equal length sequences of attributes in R and S, respectively. The IND R[X] $\subseteq$ S[Y] is satisfied in a database containing relations r over R and s over S, if for all tuples t in r, whenever the values of t are non-null then there exists a tuple u in s such that t[X] = u[Y].

   How can ISA relationships be defined in terms of inclusion dependencies ?

   How can the semantics of foreign keys be expressed in terms of inclusion dependencies ?

   [11 marks]

   [Total 33 marks]

2. Answer the following three parts.

   a. Define the concepts of primary key and entity integrity using the EMPLOYEE and DEPARTMENT relation schemas of Question 1 to illustrate your definitions.

   [11 marks]

   b. Give an incremental algorithm for testing entity integrity of a relation r over a schema R, where X $\subseteq$ schema(R) is the primary key of R.

   (Hint: Such an incremental algorithm assumes that prior to the insertion of a new tuple t into r, the relation r satisfies entity integrity. It then checks that entity integrity is maintained after t is inserted into r only with respect to t and the relation r prior to inserting t into it.)

   [11 marks]

   c. Prove (or give a convincing argument for) the assertion that a set of attributes X $\subseteq$ schema(R) is a superkey for a relation schema R if and only if for all relations r over R the cardinality of π_X(r) is the same as that of r.

   [11 marks]

[Total 33 marks]

3. A view is a relation which is defined by a relational algebra query (or equivalently an SQL query) with respect to the conceptual database. For example, we can create a view containing employee names and the names of their managers.

   Answer the following three parts.

   a. A view is **materialised** if it physically stored in the database system, or **virtual** if it is computed each time the user poses a query over it.

      What are the advantages of materialised views versus virtual views? Use the EMPLOYEE and DEPARTMENT relation schemas of Question 1 to illustrate the points.

      [11 marks]

   b. The **view update problem** is the problem of appropriately updating the conceptual database when an update is performed on a view.

      Suppose that a view is defined as a projection of a single relation onto a subset of the attributes of its relation schema, say R.

      Argue for restricting such views so that the projected attribute list contains a key for the relation schema R. Use the EMPLOYEE relation schema of Question 1 to illustrate the points.

      [11 marks]

   c. Given a fixed relation, say r, over a relation schema, R, whose attributes are PARENT and CHILD give a simple algorithm which computes the transitive closure of r.

      You are allowed to use the following relational algebra operators in your algorithm: projection, natural join, union and renaming.

      [11 marks]

   [Total 33 marks]

4. Answer the following four parts.

   a. What shortcomings of tree structures in general are B-trees and B^+-trees intended to overcome, and how do they achieve this ?

   [6 marks]

   b. What shortcomings of bulk storage of data in files is dynamic hashing intended to overcome, and how does it achieve this ?

   [5 marks]

   c. Suppose that you are a consultant on computer-based cataloguing systems. For each of the organisations below, suggest which of the three data structures mentioned in (a) and (b) you would recommend for its main system, and justify your choice.

   (i) A library for undergraduates, covering material that they need for their studies.

   (ii) A national archive (containing day-by-day collections of newspapers, government proclamations etc.) for the Central African Republic.

   (iii) A public lending library for works of light fiction.

   [12 marks]

   d. A tree-like data structure is found that has a root node with 8 pointer fields and 7 “other” fields. The remaining nodes of the tree have, 13, 11, 15, 10, 8, 10, 9 and 8 “other” fields.

   (i) Is it possible that this tree can be a B-tree? If so, what are the possible order or orders of the B-tree ? If not, explain why the information given rules out a B-tree.

   (ii) In the answer to (i) you do not need to refer to all the properties of B-trees. If you had been given a problem (d) where the answer to (i) was positive, what other checks would you make to see if the tree given was really a B-tree ?

   [10 marks]

[Total 33 marks]

5. Answer the following four parts.

   a. What is query optimisation, and why is it desirable ?

   [6 marks]

   b. With reference to any example of a particular area of application of a database that you can think of, state two queries whose evaluation will, for different reasons, not be optimal, and say why this is so.

   [6 marks]

   c. At least one lecturer, has said: “In query optimisation in database work, there is one important basic idea that you should always use as a guide when you are thinking about how to prescribe the next step in the evaluation of a complex or compound query”.

   If you agree with this, state an appropriate idea. If you disagree, give two optimising transformations that have no such idea in common. Justify your opinion.

   [11 marks]

   d. One source of knowledge about an application says that if x is an element of a relation X, then y has to be a member of some other relevant relation Y. Another source says that, if x is an element of X and z is an element of some relevant relation Z, then y must be absent from Y.

   (i) In dealing with this information, is an optimising transformation involved ? If so, why is it ? If not, why is the most suitable action not an “optimisation” ?

   (ii) How would you deal with this information ?

   [10 marks]

[Total 33 marks]

6. Answer the following three parts.

   a. If a database holds material that is interesting enough or specialised enough, it will be reasonable for someone to say that it contains knowledge. Technically, nevertheless, a database is not a knowledge base. Why ?

   [6 marks]

   b. As software systems, some knowledge bases have the form “database system plus extra components”, while other knowledge bases do not.

      (i) For those having that form, name the extra components and describe how they and the database part operate together during a computation

      (ii) Suppose that the database part is removed from your answer to (i). How is it then possible for the knowledge base to work without the help of a database ?

   [13 marks]

c. A manufacturer builds complex systems by assembling them from sub-systems, which in turn are built in the same way. At the bottom level, there is an inventory of parts that can be used in the construction of the simplest sub-systems. (If you need a concrete example, think of the construction of a car.) Any part or sub-system is known to be a part of the sub-system to which it contributes, and to have properties under the headings “dimensions” and “material”.

(i) Discuss the advantages and disadvantages of processing all this information via a database and via a knowledge base.

(ii) If a knowledge base is used, which knowledge representation seems best for the job, and why ?

(iii) If a supplier stops making some part, the manufacturer may be able to replace it by another existing part P if P has appropriate dimensions and material. Assuming that a procedure “appropriate”, with whatever arguments you want, is available to determine whether a P is suitable in this context, write pseudocode that can determine the parts that are acceptable replacements for a given part. (If your pseudocode is different from the Prolog-like notation used in lectures, explain its meaning.)

[14 marks]

[Total 33 marks]

END OF PAPER