THE UNIVERSITY
of LIVERPOOL

JANUARY 2003 EXAMINATIONS

Bachelor of Arts : Year 3
Bachelor of Science : Year 3

Topics in Databases

TIME ALLOWED : Two Hours and a half

INSTRUCTIONS TO CANDIDATES

Answer **all** questions in Section A
Answer **two** questions in Section B

If you attempt to answer more questions than the required number of questions (in any section), the marks awarded for the excess questions will be discarded (starting with your lowest mark).
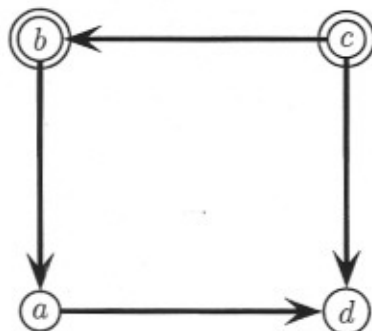
SECTION A

Answer **ALL** Questions in this section.

1. This question concerns Web-like or Semistructured Databases (theory and examples).

   (a) For the following acyclic graph G

   calculate the bracket expression for the set $(\!|v|\!)_G$ denoted by each vertex v of G (i.e., a, b, c and d). (Please, use round brackets "(" and ")" instead of the usual curly brackets "{" and "}" for denoting sets to avoid problems with distinguishing clearly the left bracket from the right one in handwritten form). **5 marks**

   (b) How can the above acyclic graph G be extended to a graph G' such that a new vertex u will denote the union of sets denoted by b and c according to the standard semantics $(\!|-|\!)$, i.e., formally, $(\!|u|\!)_{G'} = (\!|b|\!)_G \cup (\!|c|\!)_G$. **6 marks**

   (c) Give a general description of how the representation of (hyper)sets by graph vertices may be used in a (hyper)set approach to Web-like or semistructured databases. **7 marks**

   (d) Give a detailed description of how a simplified version of the WWW may be considered as a graph with labelled edges and thus may serve as an example of a Web-like Database (WDB). **8 marks**

2. This question concerns practical aspects of Semistructured Databases.

   (a) Give an example of a restructuring, rather than searching, query to the WWW which the ordinary Web search engines cannot manage. **5 marks**

   (b) Describe the *four* stages that a Web query system might use to execute general queries to the WWW or WDB, and make a conclusion comparing this with traditional search engines. **6 marks**

   (c) What do XML and HTML have in common, and what are their main differences? **5 marks**

   (d) What is the difference between the concepts of *well-formed* and *valid* XML document with respect to a document type definition DTD? What is a DTD in detail? **10 marks**

3. Questions on Deductive Databases (Datalog):

(a) In terms of a dependency graph, what is the difference between recursive and non-recursive datalog? What is the advantage of recursion? Give an example of a recursive datalog definition. **5 marks**

(b) Evaluate the following recursive program with the query

```
? - p(X,a).
```

```
IDB:
  p(X,Y) :- q(X,Y), r(X).
  q(X,Y) :- s(X,Y).
  q(X,Y) :- s(X,Z), q(Z,Y).
EDB:
  s(b,a).
  s(c,b).
  r(c).
```

5 marks

(c) What are range restricted rules in datalog? Show that each of the following rules in datalog defines an infinite relation and is not range restricted.

```
p(X,Y) :- q(X).

r(X) :- not q(X).

s(X,Y) :- X < Y.
```

6 marks

(d) Consider two relations

```
product(item, price)
```

and

```
madein(item, country).
```

The query: *"select all non-German products with price not less than 5"* can be expressed in relational algebra as

$$\pi_{1,2,4}\left(\sigma_{X_2 \geq 5 \ \& \ X_4 \neq \text{germany} \ \& \ X_1 = X_3}(\text{product} \times \text{madein})\right).$$

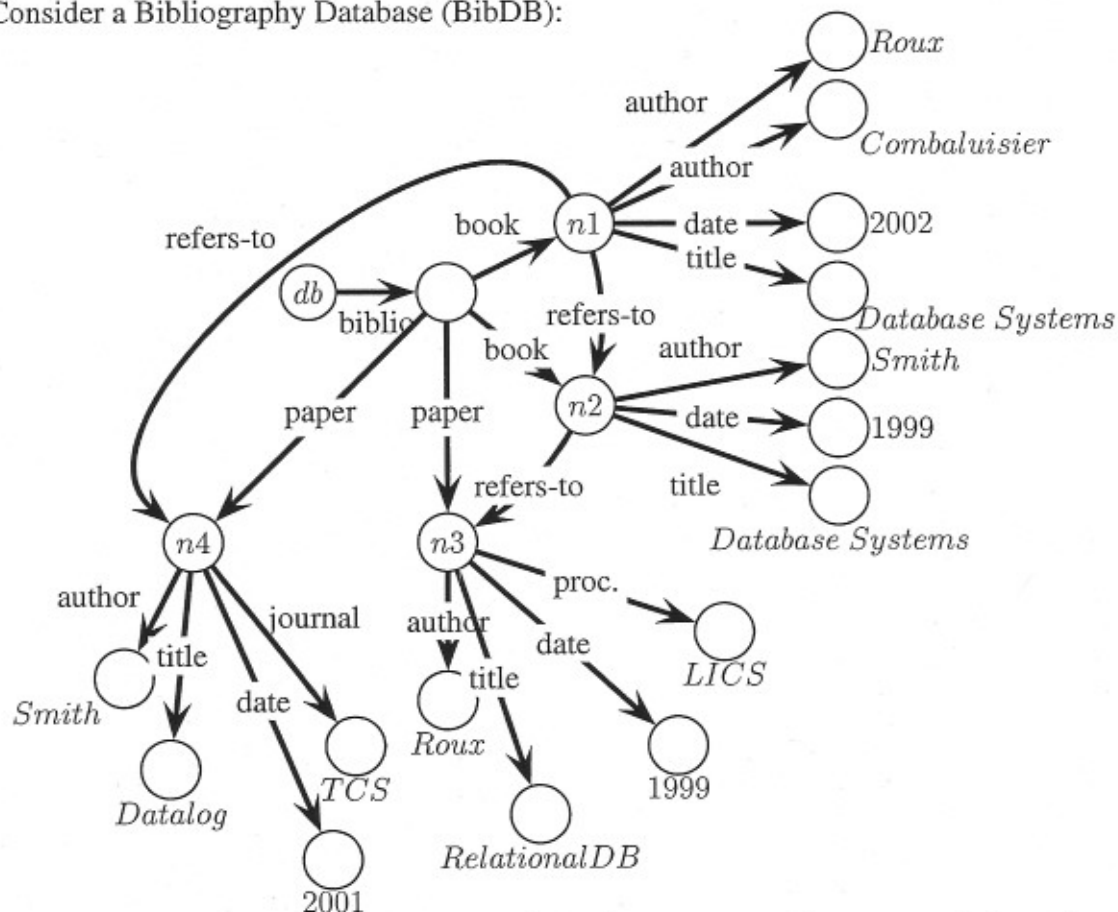Transform this query into **nr-datalog⁻**. **10 marks**

SECTION B

Answer only **TWO** Questions from this section.

Credit will be given for the best two answers in Section B.

4. Consider a Bibliography Database (BibDB):

Express, in English, the following query from the *core query language* and draw the corresponding answer graph.

```
select row: {author: A, title: T}
from biblio.book B,
     B.author A,
     B.title T,
     biblio.paper P,
     P.journal J
     B.refers-to P
```

11 marks

5. Define the syntax of the basic language $\Delta^{(*)}$ and explain its set-theoretic semantics. Which role this language plays in the set-theoretic approach to Web-like Databases. **11 marks**

6. Which extension of the language Δ^* from the previous question has a precise description of its expressive power in terms of complexity theory? Give this description and also definitions and formulations for the concepts in your answer. **11 marks**

7. Consider a bibliography database in `www.a.b.c/bib.xml` of the form

```
<bib>
   <book year="1991">
   <!-- A good introductory text -->
      <title> An introduction to Parallel Algorithms
              and Architectures
      </title>
      <author> <lastname> Leighton </lastname> </author>
      <publisher> Morgan Kaufmann </publisher>
   </book>
   <book year="1995">
      <title> Active Database Systems </title>
      <author> <lastname> Ceri </lastname> </author>
      <author> <lastname> Widom </lastname> </author>
      <publisher> Morgan Kaufmann </publisher>
   </book>
   <book year="1997">
      <title> A First Course in Database Systems </title>
      <author> <lastname> Ullman </lastname> </author>
      <author> <lastname> Widom </lastname> </author>
      <publisher> Prentice Hall </publisher>
   </book>
<\bib>
```

What is the informal meaning and the precise result of the following query in XML-QL:

```
<answer>
where <book> <author> $A </> </> in "www.a.b./bib.xml",
construct <result>
               <author> $A </>
               where <book> <author> $A </author>
                            <title>  $T </title>
                   <\book> in "www.a.b./bib.xml",
               construct <title> $T </>
           </>
</answer>
```

11 marks