Data Integration: Test #2 (due June 15, 2007)

Submit your solutions by email as a **single PDF file**. This is **individual work**. Duplicate solutions may receive reduced or no credit.

Problem 1 (30 pts)

Assume you are given two different databases, A and B, representing the same information about company sales broken down by product and year. The first database has a separate relation for every product and each of those relations has two attributes *Year* and *Amount*. The second database has one relation *Sales* with an attribute *Product* and a separate attribute for every year that contains the year's sales amount. Do not assume any fixed set of products or years.

Define XML wrappers for both databases using DTDs. Define in XQuery the mappings between the wrappers of A and B (in both directions).

Problem 2 (40 pts)

Consider the subset of XPath defined by the following grammar (where `APath` is the start symbol):

```
APath ::= Axis Rpath
RPath ::= Step  | Step Axis RPath
Axis ::= ''/'' | ''//''
Step ::= Element | Wildcard
Wildcard ::= ''*''
```

Assuming that XML documents are encoded using the following Datalog predicates:

- `root(X)` meaning `X` is a root node,
- `node(X,E)` meaning `X` is a node with tag `E`,
- `child(X,Y)` meaning `X` is a child node of the node `Y`.

Other information present in those documents is ignored.

1. Design a general method to convert any given XPath expression satisfying the above restrictions to a Datalog program that always returns (in the `query(X)` predicate the set of nodes satisfying the given XPath expression. The nodes need not have to be returned in document order. You can define arbitrary auxiliary Datalog predicates. Assume that elements are arbitrary Datalog atoms (names).
2. Show a Datalog program resulting from converting the following expression:

   `//a/*/b//*//c.`

3. Generalize your approach by allowing *conditions* on elements in the XPath expressions. The conditions may involve path expressions of the kind specified above (`RPath`) and equality comparisons but not Boolean operators. Describe the modified grammar and the extensions to the conversion method.

Problem 3 (30 pts)

(A) You are to describe the following knowledge base (TBox) using description logics:

> Each country has a (single) capital and an area (integer). Each country neighbors zero or more countries, and zero or more bodies of water. A country whose only neighbors are bodies of water is an island. Each capital is a city. Each city has a single name, lies in a single country, and has a population which is an integer.

(B) Consider the following ABox:

$bonn \in City$
$berlin \in City$
$(germany, bonn) \in HasCapital$
$(germany, berlin) \in HasCapital$
$(france, paris) \in HasCapital$

1. Is the knowledge base KB consisting of your TBox and the above ABox satisfiable?
2. Does KB imply that

   $germany \in \geq 1 HasCapital$

   $\{paris\} \sqsubseteq \forall HasCapital^{-}.\{france\}$
3. Define a class *NotCapital* consisting of cities that are not capitals.

(C) Consider the knowledge base consisting of the following Description Logic ABox:

$(eve, mark) \in Inherits.$
$(mark, jean) \in Inherits.$
$(eve, jean) \in Inherits.$
$(jean, marie) \in Inherits.$
$mark \in English.$
$marie \in \neg English.$

Does this knowledge base imply the following fact:

$eve \in (\exists Inherits.(English \sqcap \exists Inherits.\neg English))$?

Explain your answers in detail using the formal semantics of description logics.