Exercise 6 "Implementation of Databases"
Due until Thursday, 1\textsuperscript{st} June 2006 (before lecture) SS 06

6.1 Transformation by Quant Graphs

Given is the following relational database schema:

\begin{itemize}
  \item Book = \{ISBN, Author, Title, Publisher, City, IssueDate \}
  \item Reader = \{SSN, Name, City, BirthDate\}
  \item Lend = \{ISBN, SSN, Date\}
\end{itemize}

Represent the queries given below in tuple relational calculus. Build the quant graph for each of the queries. Are the query expressions benign or malicious? If a query expression is malicious, is it possible to transform this query expression into an equivalent benign query expression, using only techniques to remove redundant edges or split absorbers?

a) "Books borrowed by Miller"

b) "Books borrowed by readers living in the city where the book has been published."

c) "Readers that have borrowed a book and where exists a book that has been published after the reader has been born and borrowed a book."

6.2 Query Graphs

Given the following query: “Assistant professors managing at least one project who haven’t published a paper in 2002 and lectures, which belong to them or are held by a guest professor coming from the same city.” In tuple relational calculus, it can be expressed as:

\[\text{[ EACH prof1 IN PROFESSORS: SOME pj IN PROJECTS: (prof1.status='assistant' AND prof1.pnr= pj.pnr AND NOT SOME pa IN PAPERS (pa.year=2002 AND pa.pnr=prof1.pnr) AND SOME l in LECTURES \{(l.pnr=prof1.pnr OR SOME prof2 IN PROFESSORS (prof2.status='guest' AND prof1.city=prof2.city AND prof2.pnr=l.pnr))))]}\]

a) Construct the syntax tree of this query.
b) Is it possible to represent this query in a quant graph? Give reasons for your answer.

c) Consider the following slightly modified version of the query:

```sql
[ EACH prof1 IN PROFESSORS:
    SOME pj IN PROJECTS:
        (prof1.status='assistant' AND prof1.pnr= pj.pnr AND
        SOME pa IN PAPERS
        (pa.year=2002 AND pa.pnr=prof1.pnr) AND
        SOME l in LECTURES
        SOME prof2 IN PROFESSORS
        (prof2.status='guest' AND
        prof1.city=prof2.city AND prof2.pnr=l.pnr))]
```

Represent this query as a quant graph. Is it malicious or benign? Explain!

6.3 Query Optimization

Given are the relations r=(A,B,C), s=(C,D,E) and t=(E,F). Furthermore, the following statistics are given:

- |r| = 1000, |s|=1500, |t|=2000
- |Π_C(r)|=700, |Π_C(s)|=1100, |Π_E(s)|=50, |Π_E(t)|=250
- uniform distribution of all attribute values.

Consider the query r ⋈ s ⋈ t.

a) Describe the two possible query plans to evaluate this query.
b) Which query plan should be preferred to evaluate the query? Give reasons for your answer.