Universität Freiburg
Institut für Informatik
Prof. Dr. G. Lausen
Anas Alzoghbi
Anthony Arrascue

Georges-Köhler Allee, Geb. 51
D-79110 Freiburg
lausen@informatik.uni-freiburg.de
alzoghba@informatik.uni-freiburg.de
arrascue@informatik.uni-freiburg.de

Data Models and Query Languages
Summerterm 2014

2. Exercise Sheet: CQs, Data Integration and Chase Algorithm

Submission: 22.05.2014, 14:00
Discussion: 22.05.2014

Submission Guidlines: Please hand out your written solutions directly to your tutors right before the exercise session. If you want to submit before the deadline, you can leave your solutions in the mail box in building 51-01 (first floor). Hand written solutions are also accepted as long as these are readable.

Exercise 1 (Containment Relationships, Minimization and Set of Queries, 3 Points)
Consider the following four conjunctive queries:

- $Q_1 : ans(X, Y) \leftarrow R(X, A), R(A, B), R(B, Y), R(B, C)$
- $Q_2 : ans(X, Y) \leftarrow R(X, A), R(A, B), R(B, C), R(C, Y)$
- $Q_3 : ans(X, Y) \leftarrow R(X, Z), R(A, Y), R(X, Y)$

a) Minimize all queries **(3 pts.)**.
b) Find all equivalences and containment relationships between the above queries.
c) Consider now the following two queries:
   - $Q_1 : ans(X, Y) \leftarrow S(X, A), S(A, Y), S(X, B), S(B, Y), X < A, B < Y$
   - $Q_2 : ans(X, Y) \leftarrow S(X, Z), S(Z, Y), X < Z, Z < Y$

   Try to find a containment mapping θ as described in slide 39 from both directions (from Q_2 to Q_1 and from Q_1 to Q_2). Is it possible to assert that $Q_1 \sqsubseteq Q_2$ or $Q_2 \sqsubseteq Q_1$ based on the fact that such a mapping exists or does not exist?

Exercise 2 (Data Integration, 3 Points)
Suppose that an integration system has access to the sources S_i, for $0 < i < 4$ and A_1.

- S_1 : `SPRINGER`(<u>*ISBN*</u>, *title*, *authorID*, *suggestedRetailPrice*, *numberOfPages*)
- S_2 : `SPRINGER_AUTHOR`(<u>*authorID*</u>, *authName*)
- S_3 : `SPRINGER_COMPACT`(<u>*ISBN*</u>, *title*, *authorName*, *suggestedRetailPrice*)
- A_1 : `AMAZON_BOOK`(<u>*ISBN*</u>, *title*, *onlineID*)

The integration system requires to integrate information from both publishing houses such as SPRINGER, and online retailers such as AMAZON. Note that SPRINGER has two schemas for representing the same data: `SPRINGER` and `SPRINGER_COMPACT`.

To accomplish this the following global schema was designed:

- S_1 : BOOK(<u>*ISBN*</u>, *title*, *authorName*, *suggRetPrice*, *publisher*)
- S_2 : BOOK_SELL(<u>*ISBN*</u>, <u>*seller*</u>, *price*, *onlineID*)

In the Global As View (GAV) approach the global schema is described in terms of the local schemas whereas in the Local As View (LAV) approach the opposite direction is followed describing each local schema in terms of the global schema.

a) Provide a GAV mapping and a LAV mapping **(3 pts)**.

b) Translate the following query as a Conjunctive Query:
List the name of authors whose books are sold in Amazon for less than 35 Euro.
Reformulate your CQ with the source descriptions based on your (GAV and LAV) mappings created in a) .

c) Which advantages / disadvantages have both approaches?

Exercise 3 (Chase and Minimization, 1+2+1=4 points)
Consider the following database schema with relations

Person(*SSN*,*Name*)
Professor(*SSN*,*Name*)
Course(*CourseName*,*SSN*)
Enrolled(*CourseName*,*Participant*)

where Person stores persons including social security number (SSN) and name, Professor stores professors including social security number and name, Course contains course names and the SSN of the lecturer, and Enrolled stores course inscriptions. Further let $\Sigma := \{\alpha_1, \alpha_2, \alpha_3\}$ be the set of the following constraints.

$alpha_1 := \forall\, s, n\ (\texttt{Professor}(s,n) \rightarrow \texttt{Person}(s,n))$
$\alpha_2 := \forall\, c, s, n\ (\texttt{Course}(c,s) \wedge \texttt{Person}(s,n) \rightarrow \texttt{Professor}(s,n))$
$\alpha_3 := \forall\, c, s\ (\texttt{Course}(c,s) \rightarrow \exists\, p\ \texttt{Enrolled}(c,p))$

Further consider the Conjunctive Query

Q: $\texttt{ans}(C,N) \leftarrow \texttt{Professor}(S,N), \texttt{Course}(C,S)$

a) Describe the constraints informally **(1 pt.)**.

b) Compute Q^Σ **(2 pts.)**.

c) Compute – starting from Q^Σ – the set of all minimal Σ-equivalent queries **(1 pt.)**.