Query Processing in Data Integration Systems

Diego Calvanese

Free University of Bozen-Bolzano

BIT PhD Summer School – Bressanone
July 3–7, 2006
Structure of the course

1. Introduction to data integration
   - Basic issues in data integration
   - Logical formalization

2. Query answering in the absence of constraints
   - Global-as-view (GAV) setting
   - Local-as-view (LAV) and GLAV setting

3. Query answering in the presence of constraints
   - The role of integrity constraints
   - Global-as-view (GAV) setting
   - Local-as-view (LAV) and GLAV setting

4. Concluding remarks
Part I

Concluding remarks
Outline

1 The role of global integrity constraints

2 Query answering in GAV with constraints
   - Incompleteness and inconsistency in GAV systems
   - Query answering in GAV under inclusion dependencies
   - Rewriting CQs under inclusion dependencies in GAV
   - Query answering in GAV under IDs and KDs
   - Query answering in GAV under IDs, KDs, and EDs

3 Query answering in LAV with constraints
   - LAV systems and integrity constraints
   - Query answering in (G)LAV under inclusion dependencies
   - Query answering in (G)LAV under IDs and EDs
   - LAV systems and key dependencies
Outline

1. The role of global integrity constraints

2. Query answering in GAV with constraints
   - Incompleteness and inconsistency in GAV systems
   - Query answering in GAV under inclusion dependencies
   - Rewriting CQs under inclusion dependencies in GAV
   - Query answering in GAV under IDs and KDs
   - Query answering in GAV under IDs, KDs, and EDs

3. Query answering in LAV with constraints
   - LAV systems and integrity constraints
   - Query answering in (G)LAV under inclusion dependencies
   - Query answering in (G)LAV under IDs and EDs
   - LAV systems and key dependencies
Further issues and open problems

- Further forms of constraints, e.g.,
  - KDs with restricted forms of key-conflicting IDs
  - ontology languages, description logics, RDF
    [— & al. PODS’98, — & al., KR’06]

- Semistructured data and XML
  - constraints (DTDs, XML Schema, . . .)
  - query languages (transitive closure)

- Finite models vs. unrestricted models [Rosati, PODS’06]

- Data exchange and materialization
Acknowledgements

- Andrea Calì
- Giuseppe De Giacomo
- Domenico Lembo
- Maurizio Lenzerini
- Riccardo Rosati
- Moshe Y. Vardi
References I

S. Abiteboul and O. Duschka.
Complexity of answering queries using materialized views.

A. Calì, D. Calvanese, G. De Giacomo, and M. Lenzerini.
On the expressive power of data integration systems.

A. Calì, D. Calvanese, G. De Giacomo, and M. Lenzerini.
On the role of integrity constraints in data integration.

A. Calì, D. Lembo, and R. Rosati.
Query rewriting and answering under constraints in data integration systems.
References II

Data complexity of query answering in description logics.

D. Calvanese, G. De Giacomo, and M. Lenzerini.
On the decidability of query containment under constraints.

View-based query processing: On the relationship between rewriting, answering
and losslessness.

D. Calvanese and R. Rosati.
Answering recursive queries under keys and foreign keys is undecidable.
In Proc. of KRDB 2003. CEUR Electronic Workshop Proceedings,
Recursive query plans for data integration.

A. Fuxman and R. J. Miller.
First-order query rewriting for inconsistent databases.

G. Grahne and A. O. Mendelzon.
Tableau techniques for querying information sources through global schemas.

A. Y. Halevy.
Answering queries using views: A survey.
References IV

M. Lenzerini.
Data integration: A theoretical perspective.

Boosting information integration: The INFOMIX system.

Query answering algorithms for information agents.
R. Rosati.

On the decidability and finite controllability of query processing in databases with incomplete information.