MAGIK: Managing Completeness of Data

Ognjen Savkovic

Joint work with Werner Nutt, Sergey Paramonov and Mirza Pramita

Outline

- Introduction
- Motivation examples
- TC-QC encoding into ASP
- System architecture

Introduction

- Goal: To automate the reasoning about query completeness using completeness statements (meta-data).

- Try out the system!

  http://magik-demo.inf.unibz.it/test-version-2/

  ```
  GO: Start demo >
   Connect to Database (select school in database schema mode).
  ```

- Toy schema:
  - *pupil(name,level,code)* … *pupils*
  - *class(level,code,dept)* … *every class belongs to a department*
  - *langAtt(name,language)* … *pupils attend language courses*

Example 1: Plain reasoning

Statement 1: "We are complete for all pupils"

```
Pupil(Name,Class,Level)
```

Query 1: "Who are the pupils at the 1st class?"

```
SELECT p.name
FROM   pupil AS p
WHERE  p.level='1'
```

Question 1: Can we answer query 1 completely under the assumption of statement 1?

Statement 1: "We are complete for all pupils in the class **1a**"

```
Pupil(Name,1,a)
```

Query 1: "Who are the pupils at the 1st class?"

```
SELECT p.name
FROM   pupil AS p
WHERE  p.level='1'
```

Question 2: Can we answer query 1 completely under the assumption of statement 2?

Example 1: Plain reasoning

Statement 1: "We are complete for all pupils"

```
Pupil(Name,Class,Level)
```

Query 1: "Who are the pupils at the 1st class?"

```
SELECT p.name
FROM   pupil AS p
WHERE  p.level='1'
```

Question 1: Can we answer query 1 completely under the assumption of statement 1?

Statement 1: "We are complete for all pupils in the class **1a**"

```
Pupil(Name,1,a)
```

Query 1: "Who are the pupils at the 1st class?"

```
SELECT p.name
FROM   pupil AS p
WHERE  p.level='1'
```

Question 2: Can we answer query 1 completely under the assumption of statement 2?

Which tables are incomplete wrt Query 1?

What does MAGIK suggest to us (**Incomplete tables**)?

Example 2: Reasoning under finite domains (FD)

Statement 2: "We are complete for all pupils in the class **1a**"

```
Pupil(Name,1,a)
```

FD1: "Codes of the pupils' classes can be either 'a' or 'b'

```
pupil(code) IN {a,b}
```

Query 1: "Who are the pupils at the 1st class?"

```
SELECT p.name
FROM   pupil AS p
WHERE  p.level='1'
```

Question 3: Can we answer query 1 completely under the assumption of Statement 2 and FD1?

What is incomplete wrt *Query 1*? What MAGIK suggests to us (**Incomplete tables**)?

Example 3: Reasoning under foreign keys (FK)

Statement 3: "We are complete for French learners" `langAtt(Name,frech)`

FK1: `pupil(level,code) REFERENCES class(level,code)`

FK2: `langAtt(name) REFERENCES pupil(name)`

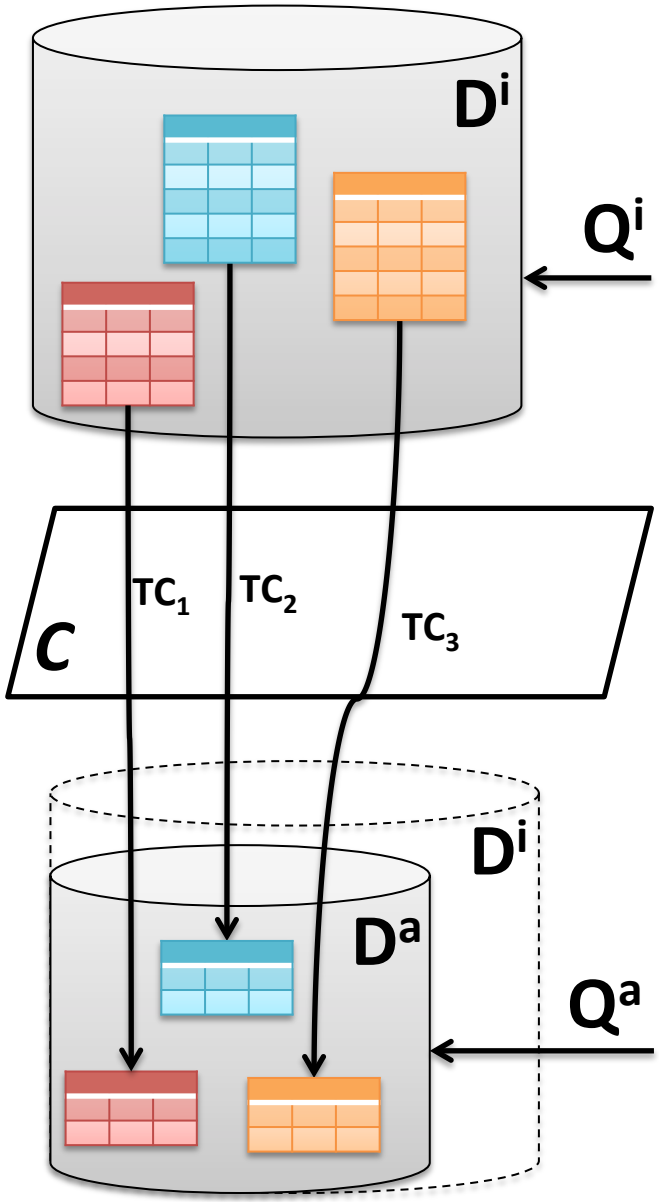
Query 2: "Which science student learns French?

```
SELECT p.name
FROM   pupila AS p, class AS c, langAtt AS l
WHERE  p.name=l.name      AND l.lang='french'
AND    p.level=c.level    AND p.code=c.code
AND    c.branch='science'
```

Question 1: Can we answer *query 2* completely under the assumption of *Statement 3*?

Question 2: Can we answer *query 2* completely under the assumption of *Statement 3* and foreign keys *FK1* and *FK2* (*enforced)?

Problem statement



A table completeness (TC) statement:

Table: *pupil(Name,Level,Code)*

Condition: *class(Level,Code,science), langAtt(Name,english)*

TC: $pupil^i(Name,Level,Code) \wedge$
$class^i(Level,Code,\ science) \wedge$
$langAtt^i(Name,english) \rightarrow pupil^a(Name,Level,Code)$

The problem statement (TC-QC): Given $TC_1,\ldots,TC_n$ and Q. Does is hold that for any partial databse (D^i,D^a) ,

If $(D^i,D^a) \models TC_1,\ldots,TC_n$ *then* $Q(D^i) = Q(D^a)$

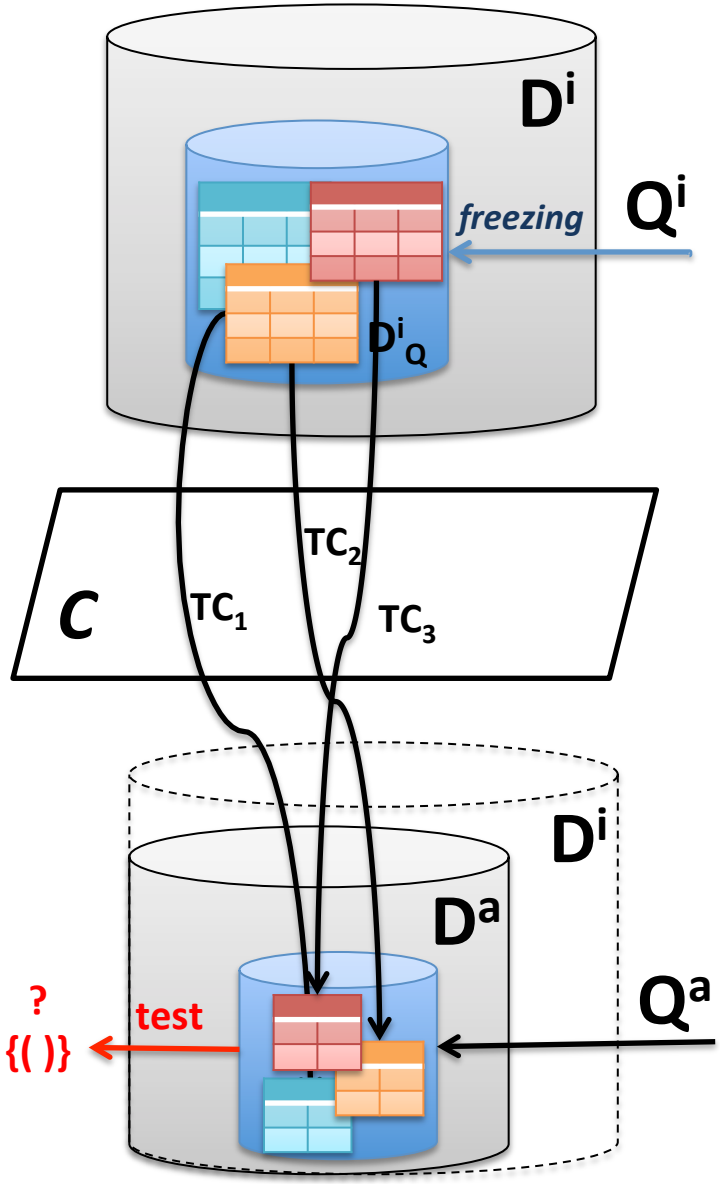
Razniewski&Nutt(VLDB'11) reduced the problem to query containment. The complexity of the problem ranges from PTIME to Π^P_2.

Our contribution:

- However they didn't show how to implement the TC-QC entailment.
- Additionally, we consider the impact of Schema Constraints, like Foreign key and Finite Domains on TC-QC entailment.

TC-QC encoding in ASP

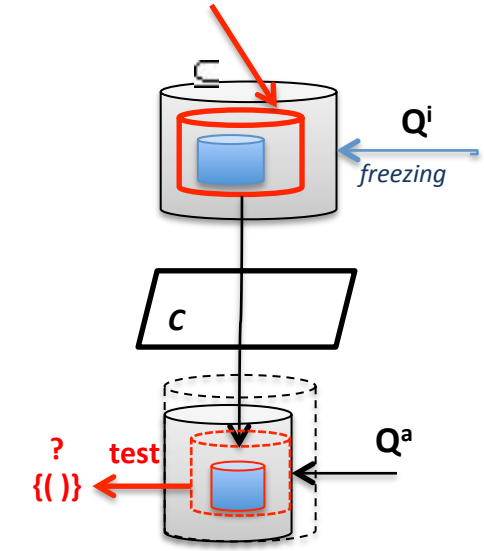
Idea: simulate the classical characterization of query containment [*Chandra and Merlin,'77*] $Q_1 \subseteq Q_2$ *iff* $Q_2(D_{Q1})$ *is not empty*
By encoding it into answer set programming (ASP).



Schema Constraints

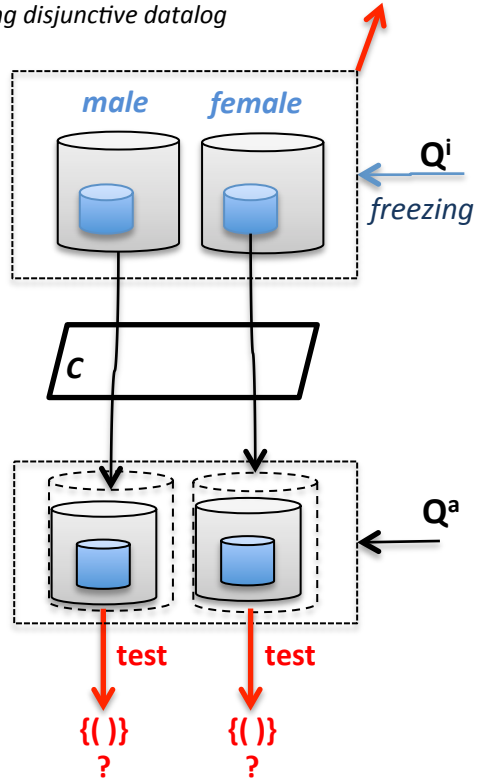
-Foreign keys *F*
[*Johnson and Klug,'84*]

Chase according F
Using Skolem terms



-Finite Domains Constraints (FDC)
f.e. *student[gender] IN {male,female}*

All possible instantiations according FDCs
Using disjunctive datalog



System Architecture