Linear Time in Ontology Mediated Querying

Marcin Przybyłko
University of Warsaw

Warsaw, 14.11.2024

University of Warsaw
Faculty
of Mathematics, Informatics
and Mechanics

Databases are incomplete

Variety	Harvest	Chamber	Date
Elise	40	0	25.09
Sabina	15	1	17.07
Witos	29	0	15.09
Helena	7	1	07.07
Natali	2	5	12.08
Wiktoria	33	4	25.07
James	12	3	20.08
Wilhelm	19	3	18.09
Regina	10	1	20.07

UNIVERSITY OF WARSAW
FACULTY
OF MATHEMATICS, INFORMATICS
AND MECHANICS

Databases are incomplete

Variety	Harvest	Chamber	Date
Elise	40	0	25.09
Sabina	15	1	17.07
Witos	29	0	15.09
Helena	7	1	07.07
Natali	2	5	12.08
Wiktoria	33	4	25.07
James	12	3	20.08
Wilhelm	19	3	18.09
Regina	10	1	20.07

- How many tons of apples were harvested this year?

University of Warsaw
Faculty of Mathematics, Informatics and Mechanics

Databases are incomplete

Variety	Harvest	Chamber	Date
Elise	40	0	25.09
Sabina	15	1	17.07
Witos	29	0	15.09
Helena	7	1	07.07
Natali	2	5	12.08
Wiktoria	33	4	25.07
James	12	3	20.08
Wilhelm	19	3	18.09
Regina	10	1	20.07

- How many tons of apples were harvested this year?

  databases are inherently incomplete

University of Warsaw
MIM
Faculty of Mathematics, Informatics and Mechanics

Databases are incomplete

Variety	Harvest	Chamber	Date
Elise	40	0	25.09
Sabina	15	1	17.07
Witos	29	0	15.09
Helena	7	1	07.07
Natali	2	5	12.08
Wiktoria	33	4	25.07
James	12	3	20.08
Wilhelm	19	3	18.09
Regina	10	1	20.07

- How many tons of apples were harvested this year?

  databases are inherently incomplete

$\text{Wilhelm} \sqcup \text{James} \sqcup \text{Witos} \sqcup \text{Elise} \sqsubseteq \text{apple}$

$(\text{Regina} \sqcup \text{Wiktoria} \sqcup \text{Natali} \sqcup \text{Helena} \sqcup \text{Sabina}) \sqcap \text{apple} \sqsubseteq \bot$

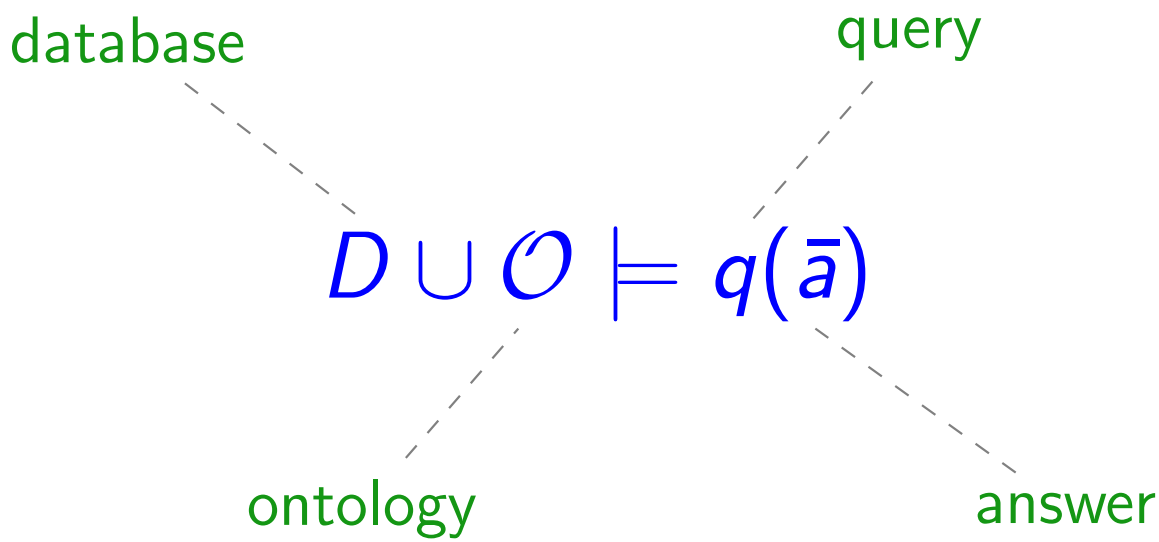
University of Warsaw
Faculty of Mathematics, Informatics and Mechanics

Ontology Mediated Querying

$$D \cup \mathcal{O} \models q(\bar{a})$$

University of Warsaw
Faculty
of Mathematics, Informatics
and Mechanics

Ontology Mediated Querying



Usual suspects:

- database D
- CQ q
- set of DL formulae $\mathcal{O}$

University of Warsaw
Faculty of Mathematics, Informatics and Mechanics

Profits

- unified interface
- more expressive queries

Challenges

- cost of reasoning
- cost of querying
- large databases, e.g. SNOMED

University of Warsaw
Faculty of Mathematics, Informatics and Mechanics

Profits

- unified interface
- more expressive queries

Challenges

- cost of reasoning
- cost of querying
- large databases, e.g. SNOMED

Task

- Separate queries that can be quickly evaluated
  - classification of $(q, \mathcal{O})$
  - linear time
  - data complexity

University of Warsaw
Faculty of Mathematics, Informatics and Mechanics

Rule of Thumb

- when rules are acyclic (aka. tree-like)
  - acyclic query $\mapsto$ linear time solutions
  - cyclic query $\mapsto$ probably no linear time algorithm

University of Warsaw
Faculty of Mathematics, Informatics and Mechanics

Rule of Thumb

- when rules are acyclic (aka. tree-like)
  - acyclic query $\mapsto$ linear time solutions
  - cyclic query $\mapsto$ probably no linear time algorithm

- upper bounds usually easy
- lower bounds surprisingly not trivial
  - involved constructions
  - conditional lower bounds

University of Warsaw
Faculty of Mathematics, Informatics and Mechanics

Future and Ongoing Work

- not acyclic rule sets
  - seems hard
    even for graphs and Datalog programs
- data constraints
  - database is complete
    rules are promises on data structure
- extensions
  - e.g., counting, partial answers, stronger rule sets

University of Warsaw
Faculty of Mathematics, Informatics and Mechanics