Advanced Database Systems (ADBS), University of Edinburgh, 2023/24

Conjunctive Queries: Minimization

(Chapter 16 of DBT)

[DBT] Database Theory, https://github.com/pdm-book/community

Complexity of Query Evaluation

Theorem: CQ-Evaluation is NP-complete, and in PTIME in data complexity

Proof:

(NP-membership) Guess-and-check:

- Consider a database D, a CQ $Q(x_1,...,x_k) := body$, and a tuple $(a_1,...,a_k)$ of values
- Guess a substitution h : terms(body) → terms(D)
- Verify that h is a match of Q in D, i.e., $h(body) \subseteq D$ and $(h(x_1),...,h(x_k)) = (a_1,...,a_k)$

(NP-hardness) Reduction from 3-colorability

(in PTIME) For every substitution h : terms(body) \rightarrow terms(D), check if h(body) \subseteq D and (h(x₁),...,h(x_k)) = (a₁,...,a_k)

Complexity of Static Analysis

Theorem: Let Q_1 and Q_2 be conjunctive queries. The problem of deciding whether there exists a query homomorphism from Q_2 to Q_1 is NP-complete

Proof:

(NP-membership) Guess a substitution, and verify that is a query homomorphism (NP-hardness) Easy reduction from CQ-Evaluation

By applying the homomorphism theorem we get that:

Corollary: CQ-Equivalence and **CQ**-Containment are NP-complete

Minimizing Conjunctive Queries

• **Goal:** minimize the number of joins in a query

- A conjunctive query Q_1 is minimal if there is no conjunctive query Q_2 such that:
 - 1. $Q_1 \equiv Q_2$
 - 2. Q_2 has fewer atoms than Q_1

 The task of CQ minimization is, given a conjunctive query Q, to compute a minimal one that is equivalent to Q

Homomorphism Theorem

A query homomorphism from $Q_1(x_1,...,x_k)$:- body₁ to $Q_2(y_1,...,y_k)$:- body₂ is a substitution h : terms(body₁) \rightarrow terms(body₂) such that:

- 1. h is a homomorphism from $body_1$ to $body_2$
- 2. $(h(x_1),...,h(x_k)) = (y_1,...,y_k)$

Homomorphism Theorem: Let Q_1 and Q_2 be conjunctive queries. It holds that:

 $Q_1 \subseteq Q_2$ iff there exists a query homomorphism from Q_2 to Q_1

Minimization by Deletion

By exploiting the homomorphism theorem we can show the following:

Theorem: Consider a conjunctive query $Q_1(x_1,...,x_k)$:- body₁.

If Q_1 is equivalent to a conjunctive query $Q_2(y_1,...,y_k)$:- body₂ where $|body_2| < |body_1|$,

then Q_1 is equivalent to a query $Q_3(x_1,...,x_k)$:- body₃ such that body₃ \subseteq body₁

∜

The above theorem says that to minimize a conjunctive query $Q_1(x_1,...,x_k)$:- body we simply need to remove some atoms from body

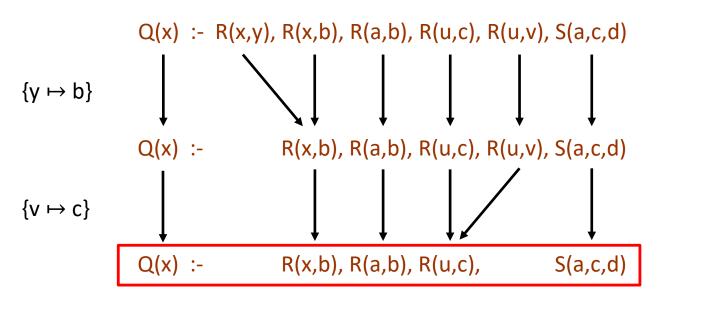
Minimization Procedure

```
\label{eq:model} \begin{split} \text{Minimization}(\mathbb{Q}(x_1,...,x_k) \ :- \ \text{body}) \\ \textbf{Repeat until no change} \\ & \text{choose an atom } \alpha \in \text{body such that the variables } x_1,...,x_k \text{ appear in body } \setminus \{\alpha\} \\ & \text{if there is a query homomorphism from } \mathbb{Q}(x_1,...,x_k) \ :- \ \text{body to } \mathbb{Q}(x_1,...,x_k) \ :- \ \text{body} \setminus \{\alpha\} \\ & \text{then body } := \ \text{body } \setminus \{\alpha\} \\ \\ \textbf{Return } \mathbb{Q}(x_1,...,x_k) \ :- \ \text{body} \end{split}
```

Note: if there is a query homomorphism from $Q(x_1,...,x_k)$:- body to $Q(x_1,...,x_k)$:- body $\setminus \{\alpha\}$, then the two queries are equivalent since there is trivially a query homomorphism from the latter to the former query

Minimization Procedure: Example

(a,b,c,d are constants)



minimal query

Note: the mapping $x \mapsto a$ is not valid since x is a distinguished variable

Uniqueness of Minimal Queries

Natural question: does the order in which we remove atoms from the body of the input conjunctive query matter?

Theorem: Consider a conjunctive query Q. Let Q_1 and Q_2 be minimal conjunctive queries such that $Q_1 \equiv Q$ and $Q_2 \equiv Q$. Then, Q_1 and Q_2 are isomorphic (i.e., they are the same up to variable renaming)

Therefore, given a conjunctive query Q, the result of Minimization(Q) is unique (up to variable renaming) and is called the core of Q