Extended Conjunctive Queries

Unions

Arithmetic

Negation

Containment of Unions of CQ's

- Theorem: $P_1 \cup \ldots \cup P_k \subseteq Q_1 \cup \ldots \cup Q_n$ if and only if for each P_i there is some Q_j such that $P_i \subseteq Q_j$.
- Proof (if): Obvious.

Proof of "Only-If"

- Assume $P_1 \cup \ldots \cup P_k \subseteq Q_1 \cup \ldots \cup Q_n$.
- Let D be the canonical (frozen) DB for P_i.
- Since the containment holds, and $P_i(D)$ includes the frozen head of P_i, there must be some Q_j such that $Q_j(D)$ also includes the frozen head of P_i.
- Thus, $P_i \subseteq Q_j$.

CQ Contained in Datalog Program

- Let Q be a CQ and P a Datalog program.
- Each returns a relation for each EDB database D, so it makes sense to ask if $Q \subseteq P$.
  - That is, $Q(D) \subseteq P(D)$ for all D.

The Containment Test

- Let D be the canonical DB for Q.
- Compute $P(D)$, and test if it contains the frozen head of Q.
- If so, $Q \subseteq P$; if not, D is a counterexample.

Example

Q: p(X,Y) :- a(X,Z) & a(Z,W) & a(W,Y)

P: p(X,Y) :- a(X,Y)

p(X,Y) :- p(X,Z) & p(Z,Y)

- Intuitively: Q = paths of length 3; P = all paths.
- Frozen Q: D = {a(x,z), a(z,w), a(w,y)}.

Example --- Continued

D = {a(x,z), a(z,w), a(w,y)}

P : p(X,Y) :- a(X,Y)

p(X,Y) :- p(X,Z) & p(Z,Y)

- Infer by first rule: p(x,z), p(z,w), p(w,y).
- Infer by second rule: p(x,w), p(z,y), p(x,y).

Frozen head of *Q*, so $Q \subseteq P$.

Other Containments

- It is doubly exponential to tell if a Datalog program is contained in a CQ.
- It is undecidable whether one Datalog program is contained in another.

CQ's With Negation

Paths of length 2 not "short-circuited."

- Allow negated subgoals.
- Example:

```
Q1: p(X,Y) :- a(X,Z) & a(Z,Y) &
                   NOT a(X,Y)
Q2: p(X,Y) :- a(X,Y) & NOT a(Y,X)
```

Unidirectional arcs.

Levy-Sagiv Test

- Test Q1 $\subseteq$ Q2 by:
  1. Consider the set of all canonical databases *D* such that the tuples of *D* are composed of only symbols 1,2,…,*n*, where *n* is the number of variables of Q1.
  2. If there is such a *D* for which Q1(*D*) $\not\subseteq$ Q2(*D*), then Q1 $\not\subseteq$ Q2.
  3. Otherwise, Q1 $\subseteq$ Q2.

Example

Q1: `p(X,Y) :- a(X,Z) & a(Z,Y) &`
`NOT a(X,Y)`

Q2: `p(X,Y) :- a(X,Y) & NOT a(Y,X)`

- Try *D* = {a(1,2), a(2,3)}.
- Q1(*D*) = {p(1,3)}.
- Q2(*D*) = {p(1,2), p(2,3)}.
- Thus, Q1 $\not\subseteq$ Q2.

Intuition

- It is not sufficient to consider only the frozen body of Q1.
- The reason is that sometimes, containment is only violated when certain variables are assigned the same constant.

CQ’s With Interpreted Predicates

- Important special case: arithmetic predicates like <.
  - A total order on values.
- General case: predicate has some specific meaning, but may not be like arithmetic comparisons.
  - Example: set-valued variables and a set-containment predicate.

CQ's With <

- To test Q1 $\subseteq$ Q2, consider all canonical DB's formed from the ordinary (not arithmetic) subgoals of Q1, by assigning each variable to one of 1,2,…,*n*.
- Equivalently: partition the variables of Q1 and order the blocks of the partition by <.

Example

Q1: p(X,Z) :- a(X,Y) & a(Y,Z) & X<Y

Q2: p(A,C) :- a(A,B) & a(B,C) & A<C

- There are 13 ordered partitions:
  - 6 orders of {X}{Y}{Z}.
  - 3*2 orders for the three 2-1 partitions, like {X}{Y,Z}.
  - 1 order for the partition {X,Y,Z}.

Example --- Continued

- Consider one ordered partition: {X,Z}{Y}; i.e., let $X = Z = 1$ and $Y = 2$.
- Then the body of

Q1: p(X,Z) :- a(X,Y) & a(Y,Z) & X<Y

becomes D ={a(1,2), a(2,1)}, and X<Y is satisfied, so the head p(1,1) is in Q1(D).

Example --- Concluded

Q2: p(A,C) :- a(A,B) & a(B,C) & A<C

D ={a(1,2), a(2,1)}

- Claim Q2(D) = ∅, since the only way to satisfy the first two subgoals are:
  1. $A = C = 1$ and $B = 2$, or
  2. $A = C = 2$ and $B = 1$.

  In either case, A<C is violated.
- Thus, Q1 $\not\subseteq$ Q2.

Arithmetic Makes Some Things Go Wrong

- Union-of-CQ's theorem no longer holds.
- Containment-mapping theorem no longer holds.

Union of CQ's With Arithmetic

P: `p(X) :- a(X) & 10≤X & X≤20`

Q: `p(X) :- a(X) & 10≤X & X≤15`

R: `p(X) :- a(X) & 15≤X & X≤20`

- $P \subseteq Q \cup R$, but neither $P \subseteq Q$ nor $P \subseteq R$ holds.

CM Theorem Doesn't Hold

Q1: `panic :- a(X,Y) & a(Y,X)`

Q2: `panic :- a(A,B) & A≤B`

- Note "panic" is a 0-ary predicate; i.e., a propositional variable.
- Q1 = "a cycle of two nodes."
- Q2 = "a nondecreasing arc."
- Notice Q1 $\subseteq$ Q2; a cycle has to be nondecreasing in one direction.

CM Theorem --- Continued

Q1: `panic :- a(X,Y) & a(Y,X)`

Q2: `panic :- a(A,B) & A≤B`

- But there is no containment mapping from Q2 to Q1, because there is no subgoal to which A≤B can be mapped.

CM Theorem for Interpreted Predicates

1. "Rectified" rules --- a normal form for CQ's with interpreted predicates.
2. A variant of the CM theorem holds for rectified rules.
   - This theorem holds for predicates other than arithmetic comparisons, but rectification uses "=" at least.

Rectification

1. No variable may appear more than once among all the argument positions of the head and all ordinary subgoals.
2. No constant may appear in the head or an ordinary subgoal.

Rectifying Rules

- Introduce new variables to replace constants or multiple occurrences of the same variable.
- Force the new variables to be equal to old variables or constants using additional equality subgoals.

Example

```
panic :- a(X,Y) & a(Y,X)
```

becomes

```
panic :- a(X,Y) & a(U,V) &
           X=V & Y=U
```

Another Example

```
p(X) :- q(X,Y,X) & r(Y,a)
```

becomes

```
p(Z) :- q(X,Y,W) & r(V,U) &
           X=W & X=Z & Y=V & U=a
```

Gupta-Zhang-Ozsoyoglu Test

- Let Q1 and Q2 be rectified rules.
- Let *M* be the set of all CM's from the ordinary (uninterpreted) subgoals of Q2 to the ordinary subgoals of Q1.
  - Note: for rectified rules, any mapping of subgoals to subgoals with the same predicate is a CM.

GZO Test --- (2)

- Theorem: Q1 $\subseteq$ Q2 if and only if the interpreted subgoals of Q1 logically imply the OR over all CM's *m* in *M* of *m* applied to the interpreted subgoals of Q2.

Example

Q2: panic :- r(X,Y) & X≤Y

Q1: panic :- r(A,B) & r(C,D) & A=D & B=C

M = {m1 , m2}

Example --- Continued

Q2: panic :- r(X,Y) & X≤Y

Q1: panic :- r(A,B) & r(C,D) & A=D & B=C

- m1(X≤Y) = A≤B; m2(X≤Y) = C≤D
- Must show:

  A=D & B=C implies (A≤B OR C≤D)

Example --- Concluded

- $A=D$ & $B=C$ implies ($A \leq B$ OR $C \leq D$)
- Proof:
  1. $A \leq B$ OR $B \leq A$ (because $<$ is a total order).
  2. $A \leq B$ OR $C \leq D$ (substitution of equals for equals).