

Rewritings for Answering Queries Using Views

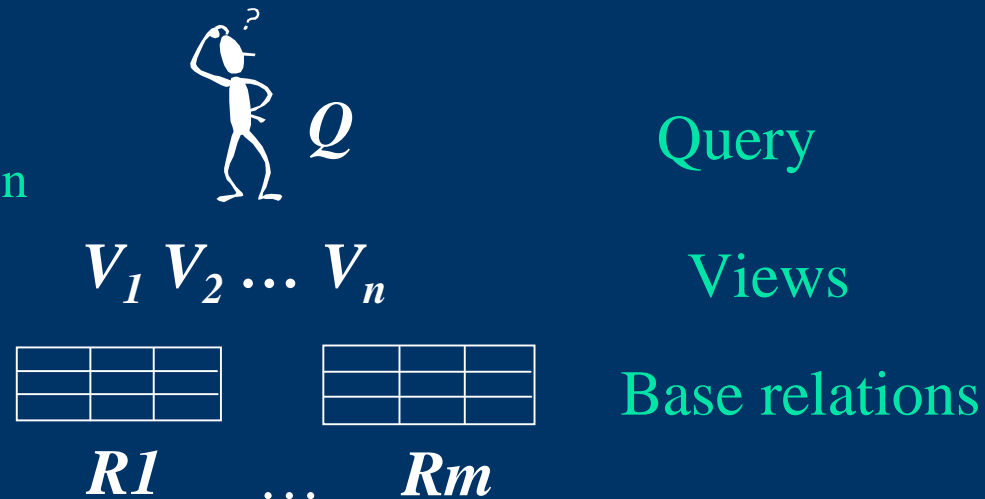


Foto Afrati

National Technical
University of Athens

Answering queries using views

- How to answer a query using *only* the results of views? [LMSS95]
- Many applications:
 - Data warehouses
 - Data integration
 - Query optimization





Information Integration Application

- Mediators represent information sources and can be thought of as views.
- Users do not know about their information sources.
- Query is asked by the user on a global schema.



Outline

Introduction [Ull97]

Maximally Contained Rewritings (MCR)

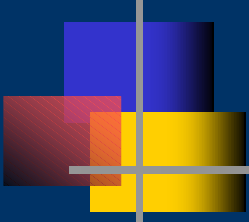
Equivalent Rewritings (ER)

CQs: Algorithms for Finding Rewritings [LMSS95,RSU95]

Efficient Plans for ERs under CWA [ALU01]

CQACs: Finding MCRs [ALM02]

Introduction [Ullman,ICDT97-TCS00]



emp (E)
phone (E,P)
office (E,O)
mgr (E,M)
dept (E,D)

Base Relations

Introduction



Views

$v1(E,P,M) :- emp(E), phone(E,P), mgr(E,M)$

$v2(E,O,D) :- emp(E), office(E,O), dept(E,D)$

$v3(E,P) :- emp(E), phone(E,P), dept(E,toy)$

Query

$Q(P,O) :- phone(sally,P), office(sally,O)$

Rewriting

$answer(P,O) :- v1(sally,P,M), v2(sally,O,D)$

Introduction



Query

$Q(P,O) :- \textit{phone}(\textit{sally},P), \textit{office}(\textit{sally},O)$

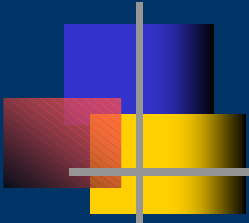
Rewriting

$\textit{answer}(P,O) :- \textit{v1}(\textit{sally},P,M), \textit{v2}(\textit{sally},O,D)$

$\textit{answer}(P,O) :- \textit{v3}(\textit{sally},P), \textit{v2}(\textit{sally},O,D)$

$\textit{answer}(P,O) \subseteq Q(P,O)$

Expansion of a rewriting



Replace the view subgoals by its definitions

answer(P,O): -*emp* (sally), *phone* (sally,P), *mgr* (sally,M),
emp (sally), *office*(sally,O), *dept* (sally,M)

answer(P,O): -*emp* (sally), *phone* (sally,P), *mgr* (sally,toy),
emp (sally), *office*(sally,O), *dept* (sally,M)

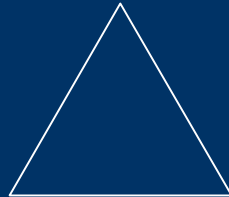
Rewriting Queries Using Views

Query Q

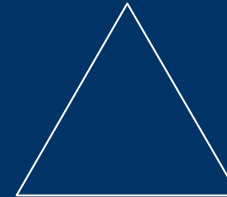
$answer () :- b1(), b2(), \dots, bn()$

Rewriting

$answer () :- v1(), v2(), \dots, vm()$



...



$answer () :- b11(), b12(), \dots, b1k() \dots bm1(), bm2(), \dots$

Expansion



Maximally Contained Rewriting

Given a query Q and view definitions V ,

P is a **maximally contained rewriting (MCR)** if,

every contained rewriting is contained in P .



Query: conjunctive query

View: conjunctive queries

Algorithms

Bound the number of subgoals in the rewriting

Without constants: Number of subgoals in a minimal contained rewriting is no more than the number of subgoals in the query. [LMSS95]

With constants: Number of subgoals in a minimal equivalent rewriting is no more than the sum of the number of subgoals and the number of variables in the query. [RSU95]

Query: conjunctive query
View: conjunctive queries

Bucket Algorithm

$Q(X, Y) :- a(X, Z), b(Z, Y)$

$v1(X, Z) :- a(X, Z), b(Z, Y)$

$v2(Z, Y) :- a(X, Z), b(Z, Y)$

$Q(X, Y) :- v1(X, Z), v2(Z, Y)$ **rewriting**



Query: conjunctive query
View: conjunctive queries

MCR is a union of conjunctive queries.

We know efficient algorithms to derive it.

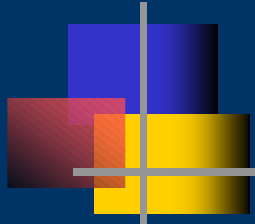
Bucket [L97]

MiniCon [PL00]

SVB [M01]

Inverse rules [DG97]

Closed/Open World Assumption

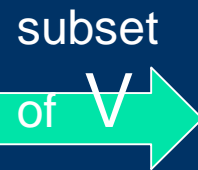


database

views



CWA



OWA



Query: conjunctive query
View: conjunctive queries

ERs under CWA

Under the CWA, all equivalent rewritings produce the same answers.

Find ERs with minimum number of subgoals

Idea:

- Find all candidate view subgoals.
- For each candidate, find what query subgoals are covered.
- Combine as in bucket.

[ALU01]



Query: conjunctive query with arithmetic comparisons

View: conjunctive queries with arithmetic comparisons

$Q(W):-r(W), W < 4$

$V1(Y,Z):-r(X), s(Y,Z), X \geq Y, Z \geq X$

$V2(Y,Z):-r(X), s(Y,Z), X \geq Y, Z > X$

Rewriting

$Q(W):-V1(W,W), W < 4$


$$Q(W):-r(W), W < 4$$

...continued

Expansion:

$$Q(W):-r(X), s(W,W), X \geq W, W \geq X, W < 4$$

Equivalent to:

$$Q(W):-r(W), s(W,W), W < 4$$



Containment of CQs with Arithmetic Comparisons

- $Q1: - Q1_0, \beta1$
- $Q2: - Q2_0, \beta2$
- $Q1 \supseteq Q2$ iff $\beta2 \Rightarrow \mu_1(\beta1) \vee \dots \vee \mu_n(\beta1)$
where μ_1, \dots, μ_n are all the containment mappings from $Q1_0$ to $Q2_0$

In $Q1_0, Q2_0$ no variable appears twice and no constant appears.



Q1: $\neg p(X,4), 4 > X$

Q2: $\neg p(X,4), p(3,X), 4 \geq X$

$Q2 \subseteq Q1$

two mappings are
needed to prove it

Q1: $\neg p(X,Y), 4 > X, Y=4$

Q2: $\neg p(X,Y), p(Z,W), 4 \geq X, Y=4, Z=3, X=W$

$\mu_1: X \rightarrow X, Y \rightarrow Y$

$\mu_2: X \rightarrow Z, Y \rightarrow W$

$4 \geq X \wedge Y=4 \wedge Z=3 \wedge X=W \Rightarrow$

$(4 > X \wedge Y=4) \vee (4 > Z \wedge W=4)$



Query: conjunctive query with Left Semi-Interval (LSI)

View: conjunctive queries with Arithmetic Comparisons(AC)

LSI : $a \geq X$

a :constant

X :variable

AC: $Y \geq X$ or $Y > X$

X : variable or constant and Y :variable

or

Y :variable or constant and X :variable



Complexity of querycontainment

CQ : NP [CM77]

CQAC : Π_2^P [K188]

CQLSI : NP [K188]

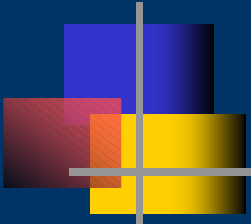
CQLSI- CQAC:NP [ALM02]



Query: conjunctive query with Left Semi-Interval (LSI)

View: conjunctive queries with Arithmetic Comparisons(AC)

- Bound the number of subgoals in the rewriting
- MCR is a union of conjunctive queries with arithmetic comparisons. [ALM02]
- Each item in the buckets is a view subgoal AND some comparison predicates.



Certain Answers

- Given: Query Q
Viewdefinitions
Viewinstance IV
- Question: Find all certain answers

Definition: A tuple t is a **certain** answer if t is an element of $Q(D)$ for any D such that $IV \subseteq V(D)$



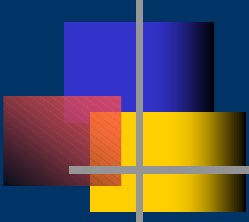
Query: conjunctive query with \neq
View: conjunctive queries

Whenever an **MCR** exists it derives
all **certain** answers. [AD98,DG98]

Deciding whether a tuple is certain answer is co-
NP complete [AD98]

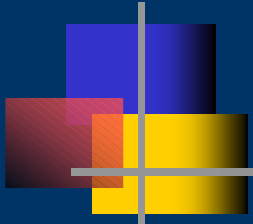
Hence there is no MCR in polynomially
evaluated languages.

Conclusion



Other interesting cases include

1. Views are unions of conjunctive queries
2. Views have binding patterns [RSU95]
3. In the presence of functional dependencies .
4. Regular Path Queries [CGLV99]
5. More on conjunctive queries with ACs [ALM02]



Thank you

