A framework for comparing query languages in their ability to express boolean queries

Dimitri Surinx [1] Jan Van den Bussche [1] Dirk Van Gucht [2]

[1]Hasselt University [2]Indiana University

Database instances

A database schema Γ is a finite nonempty set of relation names

An instance I over Γ assigns to each relation name R of arity k, a finite k-ary relation $I(R)$ over a fixed universe of data elements

The active domain $\text{adom}(I)$ of I is the set of data elements present in I

⚠ We only consider instances with a nonempty active domain

Example:

`Airports`:

id	name
1	BRU
2	BUD
3	AMS

`Flights`:

from	to
1	2
2	1
1	3
3	1

$\Gamma = \{\texttt{Airports}, \texttt{Flights}\}$

$\text{adom}(I) = \{1, 2, 3, \text{BRU}, \text{BUD}, \text{AMS}\}$

Queries

Definition

A k-ary query over a database schema Γ is a function that maps instances I over Γ to k-ary relations over $\mathrm{adom}(I)$

⚠ We require that queries are generic, i.e., $Q(f(I)) = f(Q(I))$ for any permutation f of the universe

Examples:

- Retrieve all city pairs that are two hops away from each other by plane
- Retrieve all cities pairs that are connected by plane

Boolean queries

Not all questions for a database require relational output

- Often only interested in yes/no answers

Definition

A *Boolean query* is a function that maps instance I over Γ to *true* or *false*

Examples:

- Are there two cities that are not within 4 hops of each other?
- Is the flight graph connected?
- If there is a short flight between two cities, can we then travel between the two cities by train?

Nonemptiness modality

Customary to express Boolean queries by testing nonemptiness of a query from a certain query language $\mathcal{F}$

$\rightarrow$ $Q \neq \emptyset$ is *true* on an instance I if $Q(I) \neq \emptyset$ and *false* if $Q(I) = \emptyset$

Example:

- "Are there airports with the same name" is expressed by the nonemptiness of the query

  "retrieve the different airport pairs that share the same name"

- "Are there two airports that are not connected in two hops" is expressed by the nonemptiness of the query

  "retrieve the airport pairs that are not connected in two hops"

Emptiness modality

Testing emptiness of expressions to express Boolean queries

$\rightarrow$ $Q = \emptyset$ is *true* on an instance I if $Q(I) = \emptyset$ and *false* if $Q(I) \neq \emptyset$

Examples:

- The constraint "No two airports should have the same name" is expressed by the emptiness of

  "gather the different airports with the same name"

- More generally, the FD $A \rightarrow B$ on $R(A, B)$ is expressed by the emptiness of

$$(a, b_1, b_2) \leftarrow R(a, b_1) \wedge R(a, b_2) \wedge b_1 \neq b_2$$

Containment modality

Testing containment of one expression in another one to express Boolean queries

→ $Q_1 \subseteq Q_2$ is *true* on an instance I if $Q_1(I) \subseteq Q_2(I)$ and *false* if $Q_1(I) \not\subseteq Q_2(I)$

Example:

- "If there is a short flight between two cities by plane, can we then travel the same segment by train?" is expressed by

  "retrieve the city pairs connected by direct short flights"
  $\subseteq$ "retrieve the city pairs connected by train"

- Inclusion dependencies

⚠ Gives us the ability to express a wide array of queries using weak languages

Boolean query families

We refer to nonemptiness, emptiness and containment as the *base* modalities

For any query language $\mathcal{F}$ we introduce three Boolean query families

family of Boolean queries	expressible in the form	with
$\mathcal{F}^{=\emptyset}$	$q = \emptyset$	$q \in \mathcal{F}$
$\mathcal{F}^{\neq\emptyset}$	$q \neq \emptyset$	$q \in \mathcal{F}$
$\mathcal{F}^{\subseteq}$	$q_1 \subseteq q_2$	$q_1, q_2 \in \mathcal{F}$

Framework to investigate Boolean queries

Four themes along which we can investigate Boolean queries:

1. Compare the base modalities for fixed query languages $\mathcal{F}$, e.g., $\mathcal{F}^{=\emptyset}$ vs. $\mathcal{F}^{\neq\emptyset}$
2. Compare different query languages $\mathcal{F}_1$ and $\mathcal{F}_2$ for fixed base modalities, e.g., $\mathcal{F}_1^{\subseteq}$ vs. $\mathcal{F}_2^{\subseteq}$
3. Compare different query languages $\mathcal{F}_1$ and $\mathcal{F}_2$ for different base modalities, e.g., $\mathcal{F}_1^{\subseteq}$ vs. $\mathcal{F}_2^{\neq\emptyset}$
4. Close a Boolean query family $\mathcal{B}$ under certain Boolean connectives and compare it to $\mathcal{B}$, e.g., $\mathcal{F}^{\wedge\subseteq}$ vs. $\mathcal{F}^{\subseteq}$

Note: these comparisons are uninteresting for powerful languages like FO

Theme 1

For fixed languages $\mathcal{F}$ we want to compare:

$$\mathcal{F}^{=\emptyset} \overset{?}{\subseteq} \mathcal{F}^{\neq\emptyset} \qquad \mathcal{F}^{\neq\emptyset} \overset{?}{\subseteq} \mathcal{F}^{=\emptyset}$$

$$\mathcal{F}^{\subseteq} \overset{?}{\subseteq} \mathcal{F}^{=\emptyset} \qquad \mathcal{F}^{=\emptyset} \overset{?}{\subseteq} \mathcal{F}^{\subseteq}$$

$$\mathcal{F}^{\neq\emptyset} \overset{?}{\subseteq} \mathcal{F}^{\subseteq} \qquad \mathcal{F}^{\neq\emptyset} \overset{?}{\subseteq} \mathcal{F}^{\subseteq}$$

General results:

- We identify features that enable us to go from one modality to another for fixed $\mathcal{F}$
- We identify properties that counter this ability

Applications:

- CQs and UCQs
- Navigational graph query languages

Crucial features

We identify several different query features that enable us to go from one modality to another:

- **Tests**:

$$(q_1 \text{ if } q_2)(I) = \begin{cases} q_1(I) & \text{if } q_2(I) \neq \emptyset \\ \emptyset & \text{otherwise} \end{cases}$$

- **k-ary Cylindrification**:

$$\gamma_k(q)(I) = \begin{cases} \text{adom}(I)^k & \text{if } q(I) \neq \emptyset \\ \emptyset & \text{otherwise} \end{cases}$$

- **k-ary Complementation**:

$$q^c(I) = \text{adom}(I)^k - q(I)$$

Positive results

Proposition

Let $\mathcal{F}$ be a family of queries.

1. $\mathcal{F}^{\subseteq} \subseteq \mathcal{F}^{=\emptyset}$ *if $\mathcal{F}$ is closed under set difference $(-)$.*
2. $\mathcal{F}^{=\emptyset} \subseteq \mathcal{F}^{\neq\emptyset}$ *if there exists k such that $\mathcal{F}$ is closed under*
   - *k-ary complementation, and*
   - *k-ary cylindrification.*
3. $\mathcal{F}^{\neq\emptyset} \subseteq \mathcal{F}^{\subseteq}$ *if*
   - *$\mathcal{F}$ contains a never-empty query p, and*
   - *$\mathcal{F}$ is closed under tests, or $\mathcal{F}$ is closed under k-ary cylindrification for some k.*
4. $\mathcal{F}^{=\emptyset} \subseteq \mathcal{F}^{\subseteq}$ *if $\mathcal{F}$ contains the* empty *query*

(2): $Q = \emptyset$ is equivalent to $\gamma_k(Q)^c \neq \emptyset$
(3): $Q \neq \emptyset$ is equivalent to both $p \subseteq (p \text{ if } Q)$ and $\gamma_k(p) \subseteq \gamma_k(Q)$

Negative results

Ideally, we would also like to prove that these query features are actually necessary

⚠ Cannot expect this is possible since $\mathcal{F}$ can be very pathological

Approach to solve this issue:

- Find general properties of $\mathcal{F}$ that prevent the sufficient conditions to hold

$\rightarrow$ We propose monotonicity and additivity

Monotonicity

A query Q is monotone if for any I and J we have $Q(I) \subseteq Q(I \cup J)$

$\rightarrow$ counters closure under complementation and set difference

Proposition

Let $\mathcal{F}$ be a family of monotone queries over a database schema Γ.

- *If $\mathcal{F}^{\neq\emptyset}$ contains a non-constant query, then $\mathcal{F}^{\neq\emptyset} \not\subseteq \mathcal{F}^{=\emptyset}$.*
- *If Γ contains two distinct relation names R and T of the same arity, and the two queries R and T belong to $\mathcal{F}$, then $\mathcal{F}^{\subseteq} \not\subseteq \mathcal{F}^{=\emptyset}$.*
- *If R is a binary relation name in Γ and the two queries $R \circ R$ and R belong to $\mathcal{F}$, then $\mathcal{F}^{\subseteq} \not\subseteq \mathcal{F}^{=\emptyset}$.*

Follows from: $\mathcal{F}^{=\emptyset}$ is antimonotone, $\mathcal{F}^{\neq\emptyset}$ is monotone and $\mathcal{F}^{\subseteq}$ is neither monotone nor antimonotone

Additivity

A query Q is additive if for any two active domain disjoint instances I, J we have $Q(I \cup J) = Q(I) \cup Q(J)$

$\rightarrow$ counters closure under cylindrification and test

Proposition

Let $\mathcal{F}$ be a family of additive queries.

1. *If $\mathcal{F}^{\subseteq}$ contains a non-constant query, then $\mathcal{F}^{\subseteq} \not\subseteq \mathcal{F}^{\neq\emptyset}$.*
2. *If $\mathcal{F}^{\neq\emptyset}$ contains a non-constant query, then $\mathcal{F}^{\neq\emptyset} \not\subseteq \mathcal{F}^{\subseteq}$ and $\mathcal{F}^{\neq\emptyset} \not\subseteq \mathcal{F}^{=\emptyset}$.*

Theme 1

For fixed languages $\mathcal{F}$ we want to compare:

$$\mathcal{F}^{=\emptyset} \overset{?}{\subseteq} \mathcal{F}^{\neq\emptyset} \qquad \mathcal{F}^{\neq\emptyset} \overset{?}{\subseteq} \mathcal{F}^{=\emptyset}$$

$$\mathcal{F}^{\subseteq} \overset{?}{\subseteq} \mathcal{F}^{=\emptyset} \qquad \mathcal{F}^{=\emptyset} \overset{?}{\subseteq} \mathcal{F}^{\subseteq}$$

$$\mathcal{F}^{\neq\emptyset} \overset{?}{\subseteq} \mathcal{F}^{\subseteq} \qquad \mathcal{F}^{\neq\emptyset} \overset{?}{\subseteq} \mathcal{F}^{\subseteq}$$

General results:

- We identify features that enable us to go from one modality to another for fixed $\mathcal{F}$
- We identify properties that counter this ability

Applications:

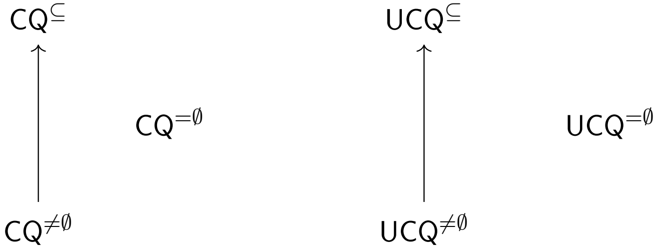
- CQs and UCQs
- Navigational graph query languages

Application: CQs and UCQs

CQs are expressions of the form $H(\overline{y}) := \exists \overline{x}\ R_1 \wedge \ldots \wedge R_n$

UCQs are unions of conjunctive queries that have a matching output schema

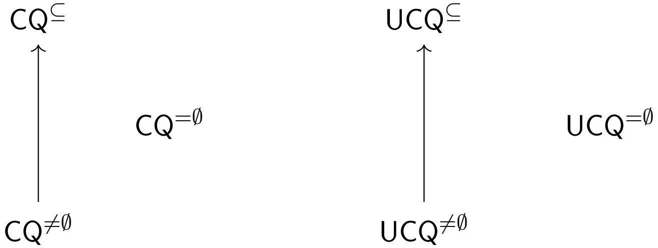
We have obtained the following:



Application: CQs and UCQs

Nearly all comparisons follow from general results since:

- Both languages are monotone
- Both languages contain a never empty query, e.g, $\exists x : x = x$
- Both are closed under tests since
  - $(\bigcup_i q_{1i} \text{ if } \bigcup_j q_{2j})$ is equivalent to the expression $\bigcup_j \bigcup_i q_{2j} \wedge q'_{1i} \neq \emptyset$ where q'_{1i} is a fully quantified version of q_{1i}



Application: Navigational graph query languages

Restriction where Γ only contains binary relation names

$\rightarrow$ can view instances over Γ as graphs

Based on the Algebra of Binary Relations [Peirce, Schröder, Tarski]

Basic features:

$$\begin{aligned}
\text{id}(G) &= \{(m, m) \mid m \in \text{adom}(G)\} \\
R(G) &= \text{the edge relation } G(R) && (\text{for } R \in \Gamma) \\
\emptyset(G) &= \emptyset \\
e_1 \circ e_2(G) &= \{(m, n) \mid \exists p : (m, p) \in e_1(G) \wedge (p, n) \in e_2(G)\} \\
e_1 \cup e_2(G) &= e_1(G) \cup e_2(G)
\end{aligned}$$

Nonbasic features

$$\mathrm{di}(G) = \{(m,n) \mid m,n \in \mathrm{adom}(G) \wedge m \neq n\}$$

$$\mathrm{all}(G) = \{(m,n) \mid m,n \in \mathrm{adom}(G)\}$$

$$e^{-1}(G) = \{(m,n) \mid (n,m) \in e(G)\}$$

$$e_1 \cap e_2(G) = e_1(G) \cap e_2(G)$$

$$e_1 - e_2(G) = e_1(G) - e_2(G)$$

$$\pi_1(e)(G) = \{(m,m) \mid \exists n : (m,n) \in e(G)\} \qquad \text{(provide tests)}$$

$$\pi_2(e)(G) = \{(m,m) \mid \exists n : (n,m) \in e(G)\}$$

$$\overline{\pi}_1(e)(G) = \{(m,m) \mid \neg\exists n : (m,n) \in e(G)\} \qquad \text{(provide negative tests)}$$

$$\overline{\pi}_2(e)(G) = \{(m,m) \mid \neg\exists n : (n,m) \in e(G)\}$$

$$e^+(G) = \text{the transitive closure of } e(G)$$

Fragments & Path Queries

A fragment is a set of features:

- Contains both projections or none of them
- Contains both coprojections or none of them
- The most basic fragment is the semiring $\{\emptyset, \mathrm{id}, \cup, \circ\}$

Let $\mathcal{N}_\Gamma(F)$ be the set of expressions built from relation names in Γ using the features in the fragment F

$\Rightarrow$ Map instances of Γ (graphs) to binary relations (*path query*)

Instead of $\mathcal{N}_\Gamma(F)^{\neq\emptyset}$, $\mathcal{N}_\Gamma(F)^{=\emptyset}$ and $\mathcal{N}_\Gamma(F)^{\subseteq}$ we write $F_\Gamma{}^{\neq\emptyset}$, $F_\Gamma{}^{=\emptyset}$ and $F_\Gamma{}^{\subseteq}$

We will omit Γ when it is not important

Examples

Consider Hobbies(person,hname), Person(pname,age)

"Output the persons that share a hobby"

- $(\text{Hobbies} \circ (\text{Hobbies}^{-1})) - \text{id}$

"Output the persons that do not have any hobbies"

- $\overline{\pi}_1(\text{Hobbies}) \circ \pi_1(\text{Person})$

Interdependencies between features

Some operators can be constructed with other operators:

$$
\begin{aligned}
\mathsf{all} &\equiv \mathsf{di} \cup \mathsf{id} \\
\mathsf{di} &\equiv \mathsf{all} - \mathsf{id} \\
e_1 \cap e_2 &\equiv e_1 - (e_1 - e_2) \\
\pi_1(e) &\equiv (e \circ e^{-1}) \cap \mathsf{id} \equiv (e \circ \mathsf{all}) \cap \mathsf{id} \equiv \overline{\pi}_1(\overline{\pi}_1(e)) \equiv \pi_2(e^{-1}) \\
\pi_2(e) &\equiv (e^{-1} \circ e) \cap \mathsf{id} \equiv (\mathsf{all} \circ e) \cap \mathsf{id} \equiv \overline{\pi}_2(\overline{\pi}_2(e)) \equiv \pi_1(e^{-1}) \\
\overline{\pi}_1(e) &\equiv \mathsf{id} - \pi_1(e) \\
\overline{\pi}_2(e) &\equiv \mathsf{id} - \pi_2(e)
\end{aligned}
$$

⚠ Operator can be expressed in $\mathcal{N}(F)$ without belonging to F

Completion of F

Define $\overline{F}$ as the smallest superset of F so that:

- if di $\in \overline{F}$, then all $\in \overline{F}$
- if all $\in \overline{F}$ and $- \in F$, then di $\in \overline{F}$
- if $- \in F$, then $\cap \in \overline{F}$
- if $\cap \in \overline{F}$ and id $\in \overline{F}$ and ($^{-1} \in F$ or all $\in \overline{F}$), then $\pi \in \overline{F}$
- if $\overline{\pi} \in \overline{F}$, then $\pi \in \overline{F}$
- if $- \in F$ and $\pi \in \overline{F}$, then $\overline{\pi} \in \overline{F}$

Example: $\overline{\{\text{all}, -\}} = \{\text{di}, \text{all}, \cap, -, \pi, \overline{\pi}\}$

Expressiveness for path queries

Well established logic: $\mathcal{N}(^{-1}, -, \mathrm{di})$ corresponds to FO^3 (Tarski & Givant)

Write $F_1 \leq F_2$ if every path query expressible in $\mathcal{N}(F_1)$ is also expressible in $\mathcal{N}(F_2)$

Theorem (Fletcher et al., 2011)

Let F_1 and F_2 be fragments. Then,

$$F_1 \leq F_2 \text{ if and only if } F_1 \subseteq \overline{F_2}$$

Theme 1 for navigational graph query fragments

Theorem

Let F be a fragment of nonbasic features. We have:

1. $F^{\subseteq} \subseteq F^{=\emptyset}$ *if and only if* $- \in F$
2. $F^{=\emptyset} \subseteq F^{\neq\emptyset}$ *if and only if* all $\in \overline{F}$ *and* ($- \in F$ *or* $\overline{\pi} \in \overline{F}$)
3. $F^{\neq\emptyset} \subseteq F^{\subseteq}$ *if and only if* all $\in \overline{F}$
4. $F^{\subseteq} \subseteq F^{\neq\emptyset}$ *if and only if* all $\in \overline{F}$ *and* $- \in F$

⚠ If and only if characterization of the proposition for general query languages $\mathcal{F}$

Proof: positive results

Theorem

Let F be a fragment of nonbasic features. We have:

1. $F^{\subseteq} \subseteq F^{=\emptyset}$ *if and only if* $- \in F$.
2. $F^{=\emptyset} \subseteq F^{\neq\emptyset}$ *if and only if* $\mathrm{all} \in \overline{F}$ *and* $(- \in F$ *or* $\overline{\pi}_1 \in \overline{F}$ *or* $\overline{\pi}_2 \in \overline{F})$.
3. $F^{\neq\emptyset} \subseteq F^{\subseteq}$ *if and only if* $\mathrm{all} \in \overline{F}$.
4. $F^{\subseteq} \subseteq F^{\neq\emptyset}$ *if and only if* $\mathrm{all} \in \overline{F}$ *and* $- \in F$.

Directly follows from the general results:

- 2-cylindrification is expressed by $\mathrm{all} \circ q \circ all$
- 1-cylindrification is expressed by $\pi_1(\mathrm{all} \circ q)$ and $\pi_2(q \circ \mathrm{all})$
- 1-complementation is expressed by $\overline{\pi}$
- 2-complementation by $\mathrm{all} - Q$

Proof: negative results

Theorem

Let F be a fragment of nonbasic features. We have:

1. $F^{\subseteq} \subseteq F^{=\emptyset}$ *if and only if* $- \in F$
2. $F^{=\emptyset} \subseteq F^{\neq\emptyset}$ *if and only if* $\text{all} \in \overline{F}$ *and* $(- \in F$ *or* $\overline{\pi}_1 \in \overline{F}$ *or* $\overline{\pi}_2 \in \overline{F})$
3. $F^{\neq\emptyset} \subseteq F^{\subseteq}$ *if and only if* $\text{all} \in \overline{F}$
4. $F^{\subseteq} \subseteq F^{\neq\emptyset}$ *if and only if* $\text{all} \in \overline{F}$ *and* $- \in F$

When $\text{all} \notin \overline{F}$

$\Rightarrow$ follows from general results since $\mathcal{N}(F)$ is additive

When $-$, $\overline{\pi}_1$ and $\overline{\pi}_2$ are not in $\overline{F}$

$\Rightarrow$ follows from general results since $\mathcal{N}(F)$ is monotone

Framework to investigate Boolean queries

Four themes along which we can investigate Boolean queries:

1. Compare the base modalities for fixed query languages $\mathcal{F}$, e.g., $\mathcal{F}^{=\emptyset}$ vs. $\mathcal{F}^{\neq\emptyset}$
2. Compare different query languages $\mathcal{F}_1$ and $\mathcal{F}_2$ for fixed base modalities, e.g., $\mathcal{F}_1^{\subseteq}$ vs. $\mathcal{F}_2^{\subseteq}$
3. Compare different query languages $\mathcal{F}_1$ and $\mathcal{F}_2$ for different base modalities, e.g., $\mathcal{F}_1^{\subseteq}$ vs. $\mathcal{F}_2^{\neq\emptyset}$
4. Close a Boolean query family $\mathcal{B}$ under certain Boolean connectives and compare it to $\mathcal{B}$, e.g., $\mathcal{F}^{\wedge\subseteq}$ vs. $\mathcal{F}^{\subseteq}$

Theme 2

For different query languages $\mathcal{F}_1$ and $\mathcal{F}_2$ we want to compare:

$$\mathcal{F}_1^{\neq\emptyset} \overset{?}{\subseteq} \mathcal{F}_2^{\neq\emptyset}$$

$$\mathcal{F}_1^{=\emptyset} \overset{?}{\subseteq} \mathcal{F}_2^{=\emptyset}$$

$$\mathcal{F}_1^{\subseteq} \overset{?}{\subseteq} F_2^{\subseteq}$$

$\rightarrow$ Particularly interesting for query languages with several query features

We focus on the navigational graph query languages

Theme 2: Nonemptiness and containment modalities

Already characterized in a larger project on the Algebra of Binary Relations

A feature g is called *primitive* for a modality $\mathcal{M} \in \{\neq\emptyset, =\emptyset, \subseteq\}$ if for every fragment F with $g \notin \overline{F}$: $\{g\}^{\mathcal{M}} \not\subseteq F^{\mathcal{M}}$

Theorem ([Fletcher et al., 2011, 2012, 2013])

The features $\mathrm{di}, \pi, \overline{\pi}, \cap$ *and* $-$ *are primitive for nonemptiness. Furthermore, for any fragment* F*:*

- $F^{\neq\emptyset} \subseteq F - \{^+\}^{\neq\emptyset}$ *iff* $F \subseteq \{\pi, \mathrm{di}, ^+\}$ *and* $|\Gamma| = 1$
- $F^{\neq\emptyset} \subseteq F - \{^{-1}\}^{\neq\emptyset}$ *iff* $^{-1} \in F$, $\pi \in \overline{F}$, $\cap \notin \overline{F}$, $^+ \notin F$
- $F^{\neq\emptyset} \subseteq F - \{\mathrm{all}\}^{\neq\emptyset}$ *iff* $F \subseteq \{\mathrm{all}, ^+\}$ *and* $|\Gamma| = 1$

Theorem (Surinx et al., 2017)

Every operator is primitive for the containment modality

Framework to investigate Boolean queries

Four themes along which we can investigate Boolean queries:

1. Compare the base modalities for fixed query languages $\mathcal{F}$, e.g., $\mathcal{F}^{=\emptyset}$ vs. $\mathcal{F}^{\neq\emptyset}$
2. Compare different query languages $\mathcal{F}_1$ and $\mathcal{F}_2$ for fixed base modalities, e.g., $\mathcal{F}_1^{\subseteq}$ vs. $\mathcal{F}_2^{\subseteq}$
3. Compare different query languages $\mathcal{F}_1$ and $\mathcal{F}_2$ for different base modalities, e.g., $\mathcal{F}_1^{\subseteq}$ vs. $\mathcal{F}_2^{\neq\emptyset}$
4. Close a Boolean query family $\mathcal{B}$ under certain Boolean connectives and compare it to $\mathcal{B}$, e.g., $\mathcal{F}^{\wedge\subseteq}$ vs. $\mathcal{F}^{\subseteq}$

Theme 3

For particular languages $\mathcal{F}_1, \mathcal{F}_2$ and modalities $\mathcal{M}_1, \mathcal{M}_2$ in $\{\neq\emptyset, =\emptyset, \subseteq\}$ we want to answer the following question:

$$\mathcal{F}_1^{\mathcal{M}_1} \overset{?}{\subseteq} \mathcal{F}_2^{\mathcal{M}_2}$$

Just as in theme 2: particularly interesting for query languages with several query features

We focus on navigational graph query languages

Reductions

For nearly all comparisons, we can directly reduce to existing results

Theorem

Let F_1 and F_2 be fragments.

1. $F_1^{\subseteq} \subseteq F_2^{=\emptyset}$ *iff* $F_1^{\subseteq} \subseteq F_2^{\subseteq}$ *and* $F_2^{\subseteq} = F_2^{=\emptyset}$
2. $F_1^{\subseteq} \subseteq F_2^{\neq\emptyset}$ *iff* $F_1^{\subseteq} \subseteq F_2^{\subseteq}$ *and* $F_2^{\subseteq} = F_2^{\neq\emptyset}$
3. $F_1^{\neq\emptyset} \subseteq F_2^{=\emptyset}$ *iff* $F_1^{\neq\emptyset} \subseteq F_2^{\neq\emptyset}$ *and* $F_2^{\neq\emptyset} = F_2^{=\emptyset}$

We still have to consider:

- $F_1^{=\emptyset} \overset{?}{\subseteq} F_2^{\subseteq}$
- $F_1^{\neq\emptyset} \overset{?}{\subseteq} F_2^{\subseteq}$

Remaining comparisons

The comparison of emptiness to containment remains open

- Harder than $F_1^{\subseteq}$ vs $F_2^{\subseteq}$ since emptiness is a special form of containment

The comparison of nonemptiness to containment also remains open

Conjecture

Let F_1 and F_2 be fragments. Then,

$$F_1^{\neq\emptyset} \subseteq F_2^{\subseteq} \text{ iff } F_1^{\neq\emptyset} \subseteq F_2^{\neq\emptyset} \text{ and } F_2^{\neq\emptyset} \subseteq F_2^{\subseteq}.$$

We prove that the conjecture holds for nearly all fragments

The open cases revolve around $F_1 = \{\pi\}$ and $F_2 \subseteq \{\mathrm{di}, {}^{-1}, {}^{+}\}$

- Proving $\{\pi\}^{\neq\emptyset} \not\subseteq \{\mathrm{di}, {}^{-1}, {}^{+}\}^{\subseteq}$ would completely prove the conjecture

An attempt to prove the open case

We focus on the union-free subfragment of $\mathcal{N}(\text{all}, ^{-1})$

- Denote this fragment with $\mathcal{A}$

We reduce to the primitivity π under the nonemptiness modality:

- $\{\pi\}^{\neq\emptyset}$ only contains monotone Boolean queries
  - $\Rightarrow$ Sufficient to look at the monotone sublanguage of $\mathcal{A}^{\subseteq}$
- Characterize the monotone sublanguage of $\mathcal{A}^{\subseteq}$ as $\mathcal{A}^{\neq\emptyset}$
  - $\rightarrow$ We prove a preservation theorem for the more general (unsafe) CQs

Monotone preservation theorem for CQs

Theorem

$\mathrm{CQ}^{\subseteq} \cap \mathrm{MON} = \mathrm{CQ}^{\neq\emptyset}$

This theorem gives a syntactical query language for a semantical sublanguage

Preservation style theorems like these are interesting in their own right

$\rightarrow$ Studied intensively in database theory, model theory and finite model theory

Monotone preservation theorem for CQs

Theorem

$\text{CQ}^{\subseteq} \cap \text{MON} = \text{CQ}^{\neq\emptyset}$

Proof:

If $Q_1 \subseteq Q_2$ is monotone:

$\Rightarrow$ $Q_1 \subseteq Q_2 \equiv \exists\text{free}(Q_1)Q_1 \subseteq \exists\text{free}(Q_2)Q_2$

$\Rightarrow$ May assume that Q_1 and Q_2 are Boolean CQs

For $Q_1 \subseteq Q_2$ where Q_1 and Q_2 are Boolean CQs we have the following trichotomy:

- $Q_1 \subseteq Q_2$ is nonmonotone, or
- $Q_1 \subseteq Q_2$ is the constant true query, or
- $Q_1 \subseteq Q_2$ is equivalent to $Q_2' \neq \emptyset$ where Q_2' equals Q_2 where some conjuncts might be removed

Framework to investigate Boolean queries

Four themes along which we can investigate Boolean queries:

1. Compare the base modalities for fixed query languages $\mathcal{F}$, e.g., $\mathcal{F}^{=\emptyset}$ vs. $\mathcal{F}^{\neq\emptyset}$
2. Compare different query languages $\mathcal{F}_1$ and $\mathcal{F}_2$ for fixed base modalities, e.g., $\mathcal{F}_1^{\subseteq}$ vs. $\mathcal{F}_2^{\subseteq}$
3. Compare different query languages $\mathcal{F}_1$ and $\mathcal{F}_2$ for different base modalities, e.g., $\mathcal{F}_1^{\subseteq}$ vs. $\mathcal{F}_2^{\neq\emptyset}$
4. Close a Boolean query family $\mathcal{B}$ under certain Boolean connectives and compare it to $\mathcal{B}$, e.g., $\mathcal{F}^{\wedge\subseteq}$ vs. $\mathcal{F}^{\subseteq}$

Theme 4: Closure properties

Comparing $\mathcal{F}^{\neq\emptyset}$ to $\mathcal{F}^{=\emptyset}$ amounts to investigating whether $\mathcal{F}^{\neq\emptyset}$ is closed under negation

We investigate closure under negation for the containment modality

Next, we generalize this idea and consider closure under conjunction

Closure under negation for the containment modality

For conjunctive queries the result is negative

Theorem

Let $\mathcal{F}$ be CQ *or* UCQ*. Then, $\mathcal{F}^{\subseteq}$ is not closed under negation.*

For navigational graph query languages we show that we need a fragment where all modalities coincide

Theorem

Let F be a fragment. Then, $F^{\subseteq}$ is closed under negation iff $- \in F$ and all $\in \overline{F}$.

Closure Under Conjunction

$CQ^{\neq\emptyset}$, $UCQ^{\neq\emptyset}$ and $UCQ^{=\emptyset}$ are trivially closed under conjunction

For $CQ^{=\emptyset}$ this changes drastically

Theorem

$CQ_\Gamma{}^{=\emptyset}$ *is closed under conjunction iff* Γ *contains at most two unary relation names and no other n-ary relation names with* $n \geq 2$

For the containment modality the result is even more strict

Theorem

$CQ_\Gamma{}^{\subseteq}$ *is closed under conjunction iff* Γ *contains at most one unary relation names and no other n-ary relation names with* $n \geq 2$

Whether $UCQ^{\subseteq}$ is closed under conjunction remains open

Closure under conjunction for graph queries

Any fragment $F^{=\emptyset}$ is closed under conjunction since union is a basic feature

For nonemptiness this changes drastically

Theorem

Let F be a fragment. Then, $F_{\Gamma}^{\neq\emptyset}$ is closed under conjunction if and only if

- *either* $\mathrm{all} \in \overline{F}$, *or*
- $|\Gamma| = 1$ *and* $F \subseteq \{^{+}\}$.

Examples:

- $R^3 \neq \emptyset \wedge R^2 \cap \mathrm{id} \neq \emptyset$ is equivalent with $R^3 \circ \mathrm{all} \circ (R^2 \cap \mathrm{id}) \neq \emptyset$
- $R^2 \circ (R^2)^+ \neq \emptyset \wedge R^7 \cup R^3 \neq \emptyset$ is equivalent with $R^4 \neq \emptyset$

Closure under conjunction for graph queries

For navigational fragments under containment we conjecture the following

Conjecture

Let F be a fragment. Then, $F^{\subseteq}$ is closed under conjunction iff $- \in F$

We have been able to prove the conjecture for:

1. $e_1 \subseteq e_2 \wedge e_3 \subseteq e_4$ is equivalent to $e_1 - e_2 \cup e_3 - e_4 = \emptyset$
2. $R^2 \subseteq R \wedge R^3 \subseteq \text{id}$ is not in $\{\text{di}, ^{-1}, ^{+}\}^{\subseteq}$
3. $R^2 \subseteq R \wedge R^3 \subseteq \emptyset$ is not in $\{\cap, \pi, ^{-1}, ^{+}\}^{\subseteq}$

(2-3) are expressible using $\overline{\pi}$
For example: $R^2 \subseteq R \wedge R^3 \subseteq \emptyset$ is equivalent to $R^2 \subseteq R \circ \overline{\pi}_1(\text{all} \circ R^3)$

Future work

Our framework can be used as a guideline to investigate Boolean queries in other contexts:

We can consider other query languages such as Codd's relational algebra

We can consider other base modalities:

- Barwise and Cooper consider the modality $e_1 \cap e_2 \neq \emptyset$ that corresponds to the natural language construct "some e_1 are e_2"
- The equality modality $e_1 = e_2$ that is true on I iff $e_1(I) = e_2(I)$

⚠ There are an infinitude of modalities one can consider. Base modalities should thus be motivated by practical use.

It would be too large of a project to provide a complete picture for all relevant Boolean query families