Classification of Annotation Semirings over Query Containment

Egor V. Kostylev    Juan L. Reutter    András Z. Salamon

LFCS, University of Edinburgh
Relational Database annotation
### Takes

<table>
<thead>
<tr>
<th>Student</th>
<th>Course</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jane</td>
<td>Algebra</td>
<td>Top Mark, Wants TA</td>
</tr>
<tr>
<td>Jane</td>
<td>Physics</td>
<td></td>
</tr>
<tr>
<td>Anne</td>
<td>History</td>
<td>Class Rep.</td>
</tr>
</tbody>
</table>

### Likes

<table>
<thead>
<tr>
<th>Student</th>
<th>Course</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jane</td>
<td>Algebra</td>
<td>Wants TA</td>
</tr>
<tr>
<td>Anne</td>
<td>Literature</td>
<td></td>
</tr>
</tbody>
</table>
### Relational Database annotation: Comments

<table>
<thead>
<tr>
<th>Takes</th>
<th>Student</th>
<th>Course</th>
<th>Likes</th>
<th>Student</th>
<th>Course</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jane</td>
<td>Algebra</td>
<td>Jane</td>
<td>Algebra</td>
<td>(Jane, Algebra):</td>
</tr>
<tr>
<td></td>
<td>Jane</td>
<td>Physics</td>
<td></td>
<td></td>
<td>{Top Mark, Wants TA}</td>
</tr>
<tr>
<td></td>
<td>Anne</td>
<td>History</td>
<td></td>
<td>Anne</td>
<td>{Class Rep.}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Anne</td>
<td>Literature</td>
<td>{Wants TA}</td>
</tr>
</tbody>
</table>

```sql
SELECT Student, Course
FROM Takes, Likes
WHERE Takes.S = Likes.S
AND Takes.C = Likes.C
```

```
(Jane, Algebra):
{Top mark, Wants TA} ∪ {Wants TA} =
{Top mark, Wants TA}
```
### Relational Database annotation: Beliefs

<table>
<thead>
<tr>
<th>Takes</th>
<th>Student</th>
<th>Course</th>
<th>{Teach. Office, Stud. Union}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jane</td>
<td>Algebra</td>
<td></td>
<td>{Teach. Office, Stud. Union}</td>
</tr>
<tr>
<td>Jane</td>
<td>Physics</td>
<td></td>
<td>{Teach. Office}</td>
</tr>
<tr>
<td>Anne</td>
<td>History</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Likes</th>
<th>Student</th>
<th>Course</th>
<th>{Stud. Union}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jane</td>
<td>Algebra</td>
<td></td>
<td>{Stud. Union}</td>
</tr>
<tr>
<td>Anne</td>
<td>Literature</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

```
SELECT Student, Course
FROM Takes, Likes
WHERE Takes.S = Likes.S
AND Takes.C = Likes.C

(Jane, Algebra):
{Teach. Office, Stud. Union} \cap
{Stud. Union} =
{Stud. Union}
```
Relational Database annotation: *Bag Semantics*

**Takes**

<table>
<thead>
<tr>
<th>Student</th>
<th>Course</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jane</td>
<td>Algebra</td>
<td>2</td>
</tr>
<tr>
<td>Jane</td>
<td>Physics</td>
<td>1</td>
</tr>
<tr>
<td>Anne</td>
<td>History</td>
<td>3</td>
</tr>
</tbody>
</table>

**Likes**

<table>
<thead>
<tr>
<th>Student</th>
<th>Course</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jane</td>
<td>Algebra</td>
<td>2</td>
</tr>
<tr>
<td>Anne</td>
<td>Literature</td>
<td>1</td>
</tr>
</tbody>
</table>

**SELECT** Student, Course  
**FROM** Takes, Likes  
**WHERE** Takes.S = Likes.S  
**AND** Takes.C = Likes.C

(Jane, Algebra):  
\[2 \times 2 = 4\]
Relational Database annotation: *Fuzzy Databases*

<table>
<thead>
<tr>
<th>Takes</th>
<th>Student</th>
<th>Course</th>
<th>Likes</th>
<th>Student</th>
<th>Course</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jane</td>
<td>Algebra</td>
<td>0.6</td>
<td>Jane</td>
<td>Algebra</td>
<td>0.5</td>
<td>0.3</td>
</tr>
<tr>
<td>Jane</td>
<td>Physics</td>
<td>0.3</td>
<td>Anne</td>
<td>Literature</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Anne</td>
<td>History</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

```
SELECT Student, Course
FROM Takes, Likes
WHERE Takes.S = Likes.S
AND Takes.C = Likes.C
(Jane, Algebra):  
0.6 \times 0.5 = 0.3
```
Semirings

(Green et al. 07):

- Domains of annotations are **commutative semirings**.
- \( \mathcal{K} = \langle K, +, \times, 0, 1 \rangle \)
- Typical example: **Natural numbers** (for Bag Semantics)
Semirings

(Green et al. 07):

- Domains of annotations are \textit{commutative semirings}.
- \( \mathcal{K} = \langle \mathcal{K}, +, \times, 0, 1 \rangle \)
- Typical example: \textbf{Natural numbers} (for \textit{Bag Semantics})

More examples:

- \textbf{Lineage} (for \textit{Comments}): \( \langle \{c_1, c_2, c_3, \cdots \}, \cup, \uplus, \emptyset, \cup \rangle \)
- \textbf{Boolean algebra} (for \textit{Beliefs}): \( \langle x, y, z, \ldots, \cup, \cap, \emptyset, \cup \rangle \)
- \textbf{Fuzzy logic} (for \textit{Fuzzy Databases}): \( \langle [0, 1], \max, \times, 0, 1 \rangle \)
Semirings

(Green et al. 07):

- Domains of annotations are **commutative semirings**.
- \( \mathcal{K} = \langle \mathbb{K}, +, \times, 0, 1 \rangle \)
- Typical example: **Natural numbers** (for **Bag Semantics**)

For query evaluation (in positive relational algebra):

- for **Joins** we **Multiply** the annotations
- for **Unions** we **Add** the annotations
Query Containment

- Optimization
- Querying using views
- Information integration
- ...

We study query containment in annotated databases
What is so **special** about query containment?

- Not always the same as for *Set Semantics*
- **Varies** depending on the annotation domain
- **Open Problems** (e.g., for *Bag Semantics*)
What is so special about query containment?

- Not always the same as for *Set Semantics*
- **Varies** depending on the annotation domain
- **Open Problems** (e.g., for *Bag Semantics*)

\[
Q_1 := \exists u \exists v \exists w \text{ Takes}(u, v), \text{ Takes}(u, w)
\]

\[
Q_2 := \exists u \exists v \text{ Takes}(u, v)
\]
What is so special about query containment?

- Not always the same as for Set Semantics
- Varies depending on the annotation domain
- Open Problems (e.g., for Bag Semantics)

\[
Q_1 := \exists u \exists v \exists w \text{ Takes}(u, v), \text{ Takes}(u, w)
\]

\[
Q_2 := \exists u \exists v \text{ Takes}(u, v)
\]

\(Q_1\) is contained in \(Q_2\) under Set Semantics

\(Q_1\) is not contained in \(Q_2\) under Bag Semantics
What is so **special** about query containment?

- Not always the same as for *Set Semantics*
- **Varies** depending on the annotation domain
- **Open Problems** (e.g., for *Bag Semantics*)

\[
\begin{align*}
Q_1 & := \exists u \exists v \exists w \text{ Takes}(u, v), \text{ Takes}(u, w) \\
Q_2 & := \exists u \exists v \text{ Takes}(u, v)
\end{align*}
\]

- **Q₂ is contained in Q₁** under *Set Semantics* or *Bag Semantics*
- **Q₂ is not contained in Q₁** over *Fuzzy Databases*
Previous Work has focused on particular semirings

- *Bag Semantics*
- *Probabilistic Databases*
- Various types of *Provenance*

But new applications may use new semirings

We focus on classes of semirings
In this talk

- Identify several classes of annotation semirings with decision procedures for checking containment of CQs.
- Generalize and structure previous work.
- Some results by known techniques (e.g., homomorphisms)
- Others using new machinery (e.g., relationships between queries and polynomials)
Outline

- Formalization of query evaluation and containment
- Known results
- Some of our results
Outline

- Formalization of query evaluation and containment
- Known results
- Some of our results
Query Evaluation on annotated databases

Bag Semantics (natural numbers): $\langle \mathbb{N}, +, \times, 0, 1 \rangle$

$Q := \exists u, \exists v, \exists w \ Takes(u, v), Takes(u, w)$

<table>
<thead>
<tr>
<th>Takes</th>
<th>Student</th>
<th>Course</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>J</td>
<td>A</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>J</td>
<td>P</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
Query Evaluation on annotated databases

*Bag Semantics* (natural numbers): $\langle \mathbb{N}, +, \times, 0, 1 \rangle$

- For each evaluation (homomorphism) $h$ from $Q$ to $I$:
  1. Compute the annotation of $h(Q)$ by multiplication,
  2. Sum over all evaluations.

$$Q := \exists u, \exists v, \exists w \text{ Takes}(u, v), \text{ Takes}(u, w)$$

<table>
<thead>
<tr>
<th>Takes</th>
<th>Student</th>
<th>Course</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$J$</td>
<td>$A$</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>$J$</td>
<td>$P$</td>
<td>1</td>
</tr>
</tbody>
</table>
Query Evaluation on annotated databases

Bag Semantics (natural numbers): \( \langle \mathbb{N}, +, \times, 0, 1 \rangle \)

- For each evaluation (homomorphism) \( h \) from \( Q \) to \( I \):
  1. Compute the annotation of \( h(Q) \) by multiplication,
  2. Sum over all evaluations.

\[
Q := \exists u, \exists v, \exists w \mbox{ Takes}(u, v), \mbox{ Takes}(u, w)
\]

\[
h(Q) := \mbox{ Takes}(J, A), \mbox{ Takes}(J, P)
\]

<table>
<thead>
<tr>
<th>Takes</th>
<th>Student</th>
<th>Course</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>( J )</td>
<td>( A )</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>( J )</td>
<td>( P )</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

\[
Q_1(I) = 2 \cdot 1
\]
Query Evaluation on annotated databases

Bag Semantics (natural numbers): \( \langle \mathbb{N}, +, \times, 0, 1 \rangle \)

- For each evaluation (homomorphism) \( h \) from \( Q \) to \( I \):
  1. Compute the annotation of \( h(Q) \) by multiplication,
  2. Sum over all evaluations.

\[
Q := \exists u, \exists v, \exists w \text{ Takes}(u, v), \text{ Takes}(u, w)
\]

\[
h(Q) := \text{ Takes}(J, P), \text{ Takes}(J, A)
\]

<table>
<thead>
<tr>
<th>Takes</th>
<th>Student</th>
<th>Course</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>J</td>
<td>A</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>J</td>
<td>P</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

\[ Q_1(I) = 2 \cdot 1 + 1 \cdot 2 \]
Query Evaluation on annotated databases

*Bag Semantics* (natural numbers): \( \langle \mathbb{N}, +, \times, 0, 1 \rangle \)

- For each evaluation (homomorphism) \( h \) from \( Q \) to \( I \):
  1. Compute the annotation of \( h(Q) \) by multiplication,
  2. Sum over all evaluations.

\[
Q := \exists u, \exists v, \exists w \text{Takes}(u, v), \text{Takes}(u, w)
\]

\[
h(Q) := \text{Takes}(J, A), \text{Takes}(J, A)
\]

<table>
<thead>
<tr>
<th>Takes</th>
<th>Student</th>
<th>Course</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>( J )</td>
<td>( A )</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>( J )</td>
<td>( P )</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

\[
Q_1(I) = 2 \cdot 1 + 1 \cdot 2 + 2 \cdot 2
\]
Bag Semantics (natural numbers): $\langle \mathbb{N}, +, \times, 0, 1 \rangle$

- For each evaluation (homomorphism) $h$ from $Q$ to $I$:
  1. Compute the annotation of $h(Q)$ by multiplication,
  2. Sum over all evaluations.

\[
Q := \exists u, \exists v, \exists w \text{ Takes}(u, v), \text{ Takes}(u, w)
\]
\[
h(Q) := \text{ Takes}(J, P), \text{ Takes}(J, P)
\]

<table>
<thead>
<tr>
<th>$I$:</th>
<th>Takes</th>
<th>Student</th>
<th>Course</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$J$</td>
<td>$A$</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>$J$</td>
<td>$P$</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

$Q_1(I) = 2 \cdot 1 + 1 \cdot 2 + 2 \cdot 2 + 1 \cdot 1 = 9$
Query Evaluation on annotated databases

Fuzzy Databases: $\langle [0, 1], \text{max}, \times, 0, 1 \rangle$

- For each evaluation $h$ from $Q$ to $I$:
  1. Compute the annotation of $h(Q)$ by multiplication,
  2. Sum over all evaluations.

$Q := \exists u, \exists v, \exists w \; \text{Takes}(u, v), \text{Takes}(u, w)$

<table>
<thead>
<tr>
<th>Takes</th>
<th>Student</th>
<th>Course</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I$:</td>
<td>$J$</td>
<td>$A$</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>$J$</td>
<td>$P$</td>
<td>0.3</td>
</tr>
</tbody>
</table>

$Q_1(I) = \max (0.7 \times 0.3, 0.3 \times 0.7, 0.7 \times 0.7, 0.3 \times 0.3) = 0.49$
Assume that the semiring $\mathcal{K}$ has a partial order $\preceq_{\mathcal{K}}$.

Definition of containment (for boolean queries):

\[ Q_1 \text{ is } \mathcal{K}\text{-contained in } Q_2 \iff Q_1(I) \preceq_{\mathcal{K}} Q_2(I), \text{ for all instances } I \]

Write $Q_1 \subseteq_{\mathcal{K}} Q_2$
Partial orders on semirings

Examples:

- For *Bag Semantics, Fuzzy Databases* the order is $\leq$,
- For *Comments, Beliefs* the order is $\subseteq$:

  $\{\text{Wants TA}\} \subseteq \{\text{Top Mark, Wants TA}\}$
Partial orders on semirings

- Examples:
  - For Bag Semantics, Fuzzy Databases the order is $\leq$.
  - For Comments, Beliefs the order is $\subseteq$:
    \[
    \{\text{Wants TA}\} \subseteq \{\text{Top Mark, Wants TA}\}
    \]
- Natural orders (defined as $a \leq_K b \iff \exists c : a + c = b$)
Partial orders on semirings

- Examples:
  - For Bag Semantics, Fuzzy Databases the order is $\leq$,
  - For Comments, Beliefs the order is $\subseteq$:
    \[ \{ \text{Wants TA} \} \subseteq \{ \text{Top Mark, Wants TA} \} \]
- Natural orders (defined as $a \preceq_K b \iff \exists c : a + c = b$)
- We are as general as possible
Partial orders on semirings

- Examples:
  - For Bag Semantics, Fuzzy Databases the order is $\leq$.
  - For Comments, Beliefs the order is $\subseteq$:
    \[ \{\text{Wants TA}\} \subseteq \{\text{Top Mark, Wants TA}\} \]

- Natural orders (defined as $a \leq_K b \iff \exists c : a + c = b$)

- We are as general as possible

- But rule out bad cases, like the semiring of Integers, for which $Q_1 \not\leq_K Q_2$ for all different $Q_1$ and $Q_2$. 
Outline

- Formalization of $\mathcal{K}$-containment
- Known results
- Some results in the paper
Previous Work

\[ \mathcal{B} = \langle \{0, 1\}, \lor, \land \rangle \ (Set \ Semantics) \]
Previous Work - Chandra & Merlin ’77

\[ Q_1 \subseteq Q_2 \text{ iff homomorphism from } Q_2 \text{ to } Q_1. \]

\( \mathcal{B} \) (Set Semantics)
Previous Work

\[ Q_1 \subseteq_K Q_2 \text{ iff homomorphism from } Q_2 \text{ to } Q_1. \]

Positive Boolean Algebra

\[ \text{PosBool (Beliefs, etc.)} \]

\[ \mathcal{B} \text{ (Set Semantics)} \]
Previous Work - Grahne ’97

Distributive Lattices

PosBool (Beliefs, etc.)

\( \mathcal{B} \) (Set Semantics)
Previous Work - Grahne '97

Distributive Lattices

\[ \mathcal{B} \ (\text{Set Semantics}) \]

\[ \text{PosBool} \ (\text{Beliefs, etc.}) \]

\[ \text{homomorphism} \]
Previous Work

$\mathcal{N}$ (Bag Semantics)

Distributive Lattices

$\mathcal{B}$ (Set Semantics)

$\text{PosBool}$ (Beliefs, etc.)

homomorphism
If surjective homomorphism from $Q_2$ to $Q_1$ then $Q_1 \subseteq_N Q_2$. 

$\mathcal{N}$ (Bag Semantics) 

Distributive Lattices 

$\mathcal{P} \mathcal{o} \mathcal{s} \mathcal{B} \mathcal{u} \mathcal{e} \mathcal{e} \mathcal{l} \mathcal{i} \mathcal{f} \ (B \mathcal{e} \mathcal{l} \mathcal{i} \mathcal{e} \mathcal{s}, \ e \mathcal{t} \mathcal{c}.)$ 

$\mathcal{B}$ (Set Semantics) 

homomorphism
If $Q_1 \subseteq_K Q_2$ then homomorphic covering from $Q_2$ to $Q_1$. 

$\mathcal{N}$ (Bag Semantics) 

Distributive Lattices 

$\mathcal{B}$ (Set Semantics) 

$\text{PosBool}$ (Beliefs, etc.) 

homomorphism
Previous Work - Green ’09

- Lineage (Comments)
  - PosBool (Beliefs, etc.)
  - \( \mathcal{B} \) (Set Semantics)
  - Distributive Lattices
    - Homomorphism
Previous Work - Green ’09

Why[X] (Why-Provenance)

Lineage (Comments)

\(\mathcal{B}\) (Set Semantics)

\(\text{PosBool}\) (Beliefs, etc.)

\(\text{Distributive Lattices}\)

homomorphism
Previous Work - Green '09

\[ \mathcal{N}[X] \text{ (Provenance Polynomials)} \]

\[ \text{Why}[X] \text{ (Why-Provenance)} \]

Polynomials over variables \( X \)

Lineage (Comments)

\[ \mathcal{B} \text{ (Set Semantics)} \]

\[ \text{PosBool} \text{ (Beliefs, etc.)} \]

Distributive Lattices

\[ \text{homomorphism} \]
Previous Work - Green ’09

\[ \mathcal{N}[X] \]

\[ \text{Why}[X] \]

\[ \text{Lineage} \]

\[ Q_1 \subseteq_{\mathcal{K}} Q_2 \text{ iff homomorphic covering from } Q_2 \text{ to } Q_1. \]

Distributive Lattices

\[ \text{PosBool} \]

\[ \mathcal{B} \]

\[ \text{homomorphism} \]
Previous Work - Green ’09

\[ \mathcal{N}[X] \]

\[ \text{Why}[X] \]

homomorphic covering

\[ \text{Lineage} \]

surjective homomorphism

Distributive Lattices

\[ \text{PosBool} \]

\[ B \]

homomorphism
Previous Work - Green '09

- **\( \mathcal{N}[X] \)**: bijective homomorphism
- **\( \text{Why}[X] \)**: surjective homomorphism
- **Lineage**: homomorphic covering
- **Distributive Lattices**
  - **PosBool**: homomorphism
  - **\( \mathcal{B} \)**
Summing up, we have:

- Different types of mappings (homomorphisms)
- For a semiring $\mathcal{K}$ they can be:
  - *Sufficient condition* for containment
  - *Necessary condition* for containment
  - *Decision procedure* for containment
Summing up, we have:

- Different types of mappings (homomorphisms)
- For a semiring $\mathcal{K}$ they can be:
  - **Sufficient condition** for containment
    
    If mapping from $Q_2$ to $Q_1$ then $Q_1 \subseteq_{\mathcal{K}} Q_2$
  
  - **Necessary condition** for containment
  - **Decision procedure** for containment
Summing up, we have:

- Different types of mappings (homomorphisms)
- For a semiring $\mathcal{K}$ they can be:
  - *Sufficient condition* for containment
  - *Necessary condition* for containment

If $Q_1 \subseteq_{\mathcal{K}} Q_2$ then mapping from $Q_2$ to $Q_1$

- *Decision procedure* for containment
Summing up, we have:

- Different types of mappings (homomorphisms)

- For a semiring $\mathcal{K}$ they can be:
  - *Sufficient condition* for containment
  - *Necessary condition* for containment
  - *Decision procedure* for containment

$Q_1 \subseteq_{\mathcal{K}} Q_2$ iff mapping from $Q_2$ to $Q_1$
We characterize the universe of semirings
We characterize the universe of semirings

- Axiomatize classes of semirings for which different types of mappings between CQs are sufficient, or necessary conditions for \( \mathcal{K} \)-containment of CQ's.
We characterize the universe of semirings

- Axiomatize classes of semirings for which different types of mappings between CQs are sufficient, or necessary conditions for \(\mathcal{K}\)-containment of CQ’s.

- The intersection of the sufficient and necessary classes for a type of mapping is a class of all semirings for which \(\mathcal{K}\)-containment is decidable by finding such a mapping.
Outline

- Formalization of $\mathcal{K}$-containment
- Known results
- Some results in the paper
  - Results for homomorphisms
  - Results for homomorphic covering
  - Results for injective homomorphisms...
    and a relevant class of polynomials
  - Results for other types of mappings
Containment of CQ's for Set Semantics

- Model Set Semantics as $B = \langle \{0, 1\}, \lor, \land, 0, 1 \rangle$

$Q_1$ is $B$-contained in $Q_2$ iff there is a homomorphism from $Q_2$ to $Q_1$
Containment of CQ's for *Set Semantics*

- Model *Set Semantics* as $B = \langle \{0, 1\}, \lor, \land, 0, 1 \rangle$

$Q_1$ is $B$-contained in $Q_2$ iff there is a homomorphism from $Q_2$ to $Q_1$

Is this true for any other semiring?
Many semirings behave as *Set Semantics*

- **PosBool** (*Event tables*),
- other boolean algebras (*Beliefs, Temporal Databases*, etc.),
- **Type A systems** (Ioannidis et al. 95),
- **Distributive lattices.**
Many semirings behave as *Set Semantics*

- **PosBool** *(Event tables)*,
- other boolean algebras *(Beliefs, Temporal Databases, etc.)*,
- Type A systems *(Ioannidis et al. 95)*,
- Distributive lattices.

Can we characterize all semirings with this behavior?
Yes we can

A semiring $\mathcal{K}$ is in $\mathcal{H}$ iff

1. $a \times a = a$
2. $1 + a = 1$

for all $a \in \mathcal{K}$. 
Yes we can

A semiring $\mathcal{K}$ is in $\mathcal{H}$ iff

1. $a \times a = a$
2. $1 + a = 1$

for all $a \in \mathcal{K}$.

Theorem

$\mathcal{H}$ captures precisely all semirings that behave as Set Semantics (w.r.t. containment of CQs)
A semiring $\mathcal{K}$ is in $\mathcal{H}$ iff

1. $a \times a = a$
2. $1 + a = 1$

for all $a \in \mathcal{K}$.

**Theorem**

$\mathcal{H}$ captures precisely all semirings that behave as Set Semantics (w.r.t. containment of CQs)

If $\mathcal{K}$ is in $\mathcal{H}$ then

- Homomorphism is a decision procedure for $\mathcal{K}$-containment
Yes we can

A semiring $\mathcal{K}$ is in $\mathcal{H}$ iff
1. $a \times a = a$
2. $1 + a = 1$
for all $a \in \mathcal{K}$.

Theorem

$\mathcal{H}$ captures precisely all semirings that behave as Set Semantics (w.r.t. containment of CQs)

If Homomorphism is a decision procedure for $\mathcal{K}$-containment
  ▶ Then $\mathcal{K}$ is in $\mathcal{H}$
Class $\mathcal{H}$ for homomorphisms

- $\mathcal{N}[X]$
- Why$[X]$
- $\mathcal{N}$
- Lineage

**Homomorphism from** $Q_2$ **to** $Q_1$, **iff** $Q_1 \subseteq_K Q_2$

Distributive Lattices

- $\text{PosBool}$
- $B$

$a \times a = a$
$1 + a = 1$
Outline

- Formalization of $K$-containment
- Known results
- Some results in the paper
  - Results for homomorphisms
  - Results for homomorphic covering
  - Results for injective homomorphisms...
    and a relevant class of polynomials
  - Results for other types of mappings
Moving away from $\mathcal{H}$

Two options:

- Keep $a \times a = a$
- Keep $1 + a = 1$
Moving away from $\mathcal{H}$

Two options:

- Keep $a \times a = a$
- Keep $1 + a = 1$

Example:

- Semiring for Comments

  $\text{Lineage} = \langle \{x, y, z, w, \ldots \}, \cup, \uplus \rangle$
Semirings satisfying $a \times a = a$
Semirings satisfying $a \times a = a$

- Homomorphisms are not sufficient condition

$$Q_1 := \exists u \exists v \exists w \ Takes(u, v), \ Likes(u, w)$$
$$Q_2 := \exists u \exists v \ Takes(u, v)$$

- There exists a Homomorphism from $Q_2$ to $Q_1$
- $Q_1$ is not Lineage-contained in $Q_2$
Semirings satisfying $a \times a = a$

- Homomorphisms are not sufficient condition

$Q_1 := \exists u \exists v \exists w \ Takes(u, v), Likes(u, w)$

$Q_2 := \exists u \exists v \ Takes(u, v)$

<table>
<thead>
<tr>
<th>Takes</th>
<th>Student</th>
<th>Course</th>
<th>Lineage</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I$:</td>
<td>$J$</td>
<td>$A$</td>
<td>$x$</td>
</tr>
<tr>
<td></td>
<td>$J$</td>
<td>$P$</td>
<td>$x$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Likes</th>
<th>Student</th>
<th>Course</th>
<th>Lineage</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J$</td>
<td>$A$</td>
<td></td>
<td>$y$</td>
</tr>
</tbody>
</table>
Semirings satisfying \( a \times a = a \)

- Homomorphisms are not sufficient condition

\[
Q_1 := \exists u \exists v \exists w \text{ Takes}(u, v), \text{ Likes}(u, w) \\
Q_2 := \exists u \exists v \text{ Takes}(u, v)
\]

<table>
<thead>
<tr>
<th>Takes</th>
<th>Student</th>
<th>Course</th>
<th>Lineage</th>
</tr>
</thead>
<tbody>
<tr>
<td>( I ):</td>
<td>J</td>
<td>A</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>P</td>
<td>x</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Likes</th>
<th>Student</th>
<th>Course</th>
<th>Lineage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>J</td>
<td>A</td>
<td>y</td>
</tr>
</tbody>
</table>

- \( Q_1(I) = \{x, y\} \)
- \( Q_2(I) = \{x\} \)
We need a stricter notion of mapping

Idea:

- force both queries to target the same relations
Homomorphic Covering from $Q_1$ to $Q_2$

Intuition:
Cover each atom of $Q_2$ with a homomorphism from $Q_1$ to $Q_2$

\[
Q_1 := \exists u \exists v \exists w \text{ Takes}(u, v), \text{ Likes}(u, w)
\]
\[
Q_2 := \exists u \exists v \exists w \text{ Takes}(u, v), \text{ Takes}(u, w), \text{ Likes}(u, w)
\]
Homomorphic Covering from $Q_1$ to $Q_2$

Intuition:
Cover each atom of $Q_2$ with a homomorphism from $Q_1$ to $Q_2$

$Q_1 := \exists u \exists v \exists w \ Takes(u, v), Likes(u, w)$
$Q_2 := \exists u \exists v \exists w \ Takes(u, v), Takes(u, w), Likes(u, w)$
Homomorphic Covering from $Q_1$ to $Q_2$

Intuition:
Cover each atom of $Q_2$ with a homomorphism from $Q_1$ to $Q_2$

\[
\begin{align*}
Q_1 & : = \exists u \exists v \exists w \ \text{Takes}(u, v), \ \text{Likes}(u, w) \\
Q_2 & : = \exists u \exists v \exists w \ \text{Takes}(u, v), \ \text{Takes}(u, w), \ \text{Likes}(u, w)
\end{align*}
\]
Homomorphic Covering from $Q_1$ to $Q_2$

Intuition:

Cover each atom of $Q_2$ with a homomorphism from $Q_1$ to $Q_2$

\[
Q_1 := \exists u \exists v \exists w \text{ Takes}(u, v), \text{ Likes}(u, w)
\]

\[
Q_2 := \exists u \exists v \exists w \text{ Takes}(u, v), \text{ Takes}(u, w), \text{ Likes}(u, w)
\]
Homomorphic Covering from $Q_1$ to $Q_2$

Intuition:
Cover each atom of $Q_2$ with a homomorphism from $Q_1$ to $Q_2$

$Q_1 := \exists u \exists v \exists w \text{ Takes}(u, v), \text{ Likes}(u, w)$

$Q_2 := \exists u \exists v \exists w \text{ Takes}(u, v), \text{ Takes}(u, w), \text{ Likes}(u, w)$
Homomorphic Covering from $Q_1$ to $Q_2$

Intuition:
Cover each atom of $Q_2$ with a homomorphism from $Q_1$ to $Q_2$

\[ Q_1 := \exists u \exists v \exists w \; \text{Takes}(u, v), \text{Likes}(u, w) \]
\[ Q_2 := \exists u \exists v \exists w \; \text{Takes}(u, v), \text{Takes}(u, w), \text{Likes}(u, w) \]

There is a homomorphic covering from $Q_1$ to $Q_2$
Homomorphic Covering from $Q_1$ to $Q_2$

Intuition:
Cover each atom of $Q_2$ with a homomorphism from $Q_1$ to $Q_2$

$Q_1 := \exists u \exists v \exists w \ Takes(u, v), Likes(u, w)$
$Q_2 := \exists u \exists v \exists w \ Takes(u, v), Takes(u, w), Likes(u, w)$

There is a homomorphistic covering from $Q_1$ to $Q_2$

$Q_3 := \exists u \exists v \ Takes(u, v)$
$Q_4 := \exists u \exists v \exists w \ Takes(u, v), Likes(u, w)$

There is no homomorphistic covering from $Q_3$ to $Q_4$
We can now capture semirings satisfying $a \times a = a$

Let $\mathcal{K}$ be a semiring.

Theorem

If $\mathcal{K}$ satisfies $a \times a = a$

- Then Homomorphic covering is a sufficient condition for $\mathcal{K}$-containment
We can now capture semirings satisfying $a \times a = a$

Let $\mathcal{K}$ be a semiring.

Theorem

If $\mathcal{K}$ satisfies $a \times a = a$

- Then Homomorphic covering is a sufficient condition for $\mathcal{K}$-containment

If Homomorphic covering is a sufficient condition for $\mathcal{K}$-containment

- Then, $\mathcal{K}$ satisfies $a \times a = a$
We can now capture semirings satisfying \( a \times a = a \)

Let \( \mathcal{K} \) be a semiring.

**Theorem**

If \( \mathcal{K} \) satisfies \( a \times a = a \)

- Then *Homomorphic covering is a sufficient condition for \( \mathcal{K} \)-containment*

If *Homomorphic covering is a sufficient condition for \( \mathcal{K} \)-containment*

- Then, \( \mathcal{K} \) satisfies \( a \times a = a \)

\( a \times a = a \) captures homomorphic covering, as a *sufficient condition*. 
Class for $a \times a = a$

If homomorphic covering from $Q_2$ to $Q_1$, then $Q_1 \subseteq_{K} Q_2$.
How to describe semirings $\mathcal{K}$ for which homomorphic covering is a necessary condition for $\mathcal{K}$-containment?
Homomorphic covering as a necessary condition

How to describe semirings $\mathcal{K}$ for which homomorphic covering is a necessary condition for $\mathcal{K}$-containment?

Use the following axiom:

$$\Phi : \forall n, k \geq 1, \forall x_1, \ldots, x_n, y$$

$$x_1 \times \ldots \times x_n \times y \not\preceq_{\mathcal{K}} (x_1 + \ldots + x_n)^k$$
Homomorphic covering as a necessary condition

How to describe semirings \( \mathcal{K} \) for which homomorphic covering is a necessary condition for \( \mathcal{K} \)-containment?

Use the following axiom:

\[
\Phi : \forall n, k \geq 1, \ \forall x_1, \ldots, x_n, y \quad x_1 \times \ldots \times x_n \times y \not\preceq_{\mathcal{K}} (x_1 + \ldots + x_n)^k
\]

Note that \( \Phi \) holds for Bag Semantics, but not for Set Semantics.
Homomorphic covering as a necessary condition

Theorem

If $\mathcal{K}$ satisfies $\Phi$

- then Homomorphic covering is a necessary condition for $\mathcal{K}$-containment
Homomorphic covering as a necessary condition

Theorem

If $\mathcal{K}$ satisfies $\Phi$,

- then **Homomorphic covering is a necessary condition for $\mathcal{K}$-containment**

If **Homomorphic covering is a necessary condition for $\mathcal{K}$-containment**

- then, $\mathcal{K}$ satisfies $\Phi$
Homomorphic covering as a necessary condition

Theorem

If $\mathcal{K}$ satisfies $\Phi$

- then Homomorphic covering is a necessary condition for $\mathcal{K}$-containment

If Homomorphic covering is a necessary condition for $\mathcal{K}$-containment

- then, $\mathcal{K}$ satisfies $\Phi$

$\Phi$ captures homomorphic covering, as a necessary condition.
Obtain a class where containment is decidable

Theorem

The following are equivalent:

- $\mathcal{K}$ satisfies $a \times a = a$ and $\Phi$
- $Q_1 \subseteq_{\mathcal{K}} Q_2$ iff homomorphict covering from $Q_2$ to $Q_1$

Gives us a large class of semirings where $\mathcal{K}$-containment is decidable by a simple procedure.
Obtain a class where containment is decidable

Class $\Phi$  

Why $[X]$  

$\mathcal{N}[X]$  

Lineage  

$\mathcal{N}$  

Homomorphic covering from $Q_2$ to $Q_1$, iff $Q_1 \subseteq_K Q_2$  

$a \times a = a$
Outline

- Formalization of $\mathcal{K}$-containment
- Known results
- Some results in the paper
  - Results for homomorphisms
  - Results for homomorphic covering
  - Results for injective homomorphisms...
    and a relevant class of polynomials
  - Results for other types of mappings
Moving away from $\mathcal{H}$

Two options:

- Keep $a \times a = a$
- Keep $1 + a = 1$
Injective homomorphisms

A homomorphism from $Q_2$ to $Q_1$ is injective iff different atoms of $Q_2$ are mapped into different atoms of $Q_1$. 
We can now capture semirings satisfying $1 + a = 1$.

Let $\mathcal{K}$ be a semiring.

**Theorem**

*If $\mathcal{K}$ satisfies $1 + a = 1$ then injective homomorphism is a sufficient condition for $\mathcal{K}$-containment.*

*If injective homomorphism is a sufficient condition for $\mathcal{K}$-containment then, $\mathcal{K}$ satisfies $1 + a = 1$.*

$1 + a = 1$ captures injective homomorphism, as a sufficient condition.
Injective homomorphism as a necessary condition

How to describe semirings $\mathcal{K}$ for which injective homomorphism is a necessary condition for $\mathcal{K}$-containment?
Injective homomorphism as a necessary condition

How to describe semirings $\mathcal{K}$ for which injective homomorphism is a necessary condition for $\mathcal{K}$-containment?

Even more elaborate than the axiom $\Phi$ – we need a notion of CQ-admissible polynomials.
CQ-admissible polynomials

- The semiring of *Provenance Polynomials*

\[ \mathcal{N}[X] = \langle \mathbb{N}[X], +, \times, 0, 1 \rangle \]

is the most general (free) of commutative semirings (Green et al. 2007).
CQ-admissible polynomials

- The semiring of Provenance Polynomials

\[ \mathcal{N}[X] = \langle \mathbb{N}[X], +, \times, 0, 1 \rangle \]

is the most general (free) of commutative semirings (Green et al. 2007).

- A database instance \( I \) is abstractly tagged if every tuple is annotated by a unique variable from \( X \).
A polynomial $P$ is \textit{CQ-admissible} iff there exists a CQ $Q$ and an abstractly tagged instance $I$ such that $P$ evaluates to $Q(I)$ over $\mathcal{N}[X]$. 
Example of a CQ-admissible polynomial

\[ Q_1 := \exists u \exists v \exists w \text{ Takes}(u, v), \text{ Takes}(u, w) \]

<table>
<thead>
<tr>
<th>Takes</th>
<th>Student</th>
<th>Course</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( I ):</td>
<td>( J )</td>
<td>( A )</td>
<td>( x )</td>
</tr>
<tr>
<td>( J )</td>
<td>( P )</td>
<td>( y )</td>
<td></td>
</tr>
</tbody>
</table>

\[ Q_1(I) = x \cdot y + y \cdot x + x \cdot x + y \cdot y = x^2 + 2xy + y^2 \]
CQ-admissible polynomials

- **Remark**: factorized by Dan and Jakub!
CQ-admissible polynomials

- **Remark:** factorized by Dan and Jakub!
- We need to understand the **structure** of these polynomials.
CQ-admissible polynomials

Some properties:

- Not every polynomial is CQ-admissible:
  - Only homogeneous polynomials are (the degree is the number of atoms in the query)
CQ-admissible polynomials

Some properties:

- Not every polynomial is CQ-admissible:
  - Only homogeneous polynomials are (the degree is the number of atoms in the query)
  - Not every homogeneous is (e.g., $x^2 + xy + y^2$ is not CQ-admissible)
CQ-admissible polynomials

Some properties:

- Not every polynomial is CQ-admissible:
  - Only homogeneous polynomials are
    (the degree is the number of atoms in the query)
  - Not every homogeneous is
    (e.g., $x^2 + xy + y^2$ is not CQ-admissible)

- In the paper we give a syntactic characterization of
  CQ-admissible polynomials (not that elegant, really).
CQ-admissible polynomials

Some properties:

- Not every polynomial is CQ-admissible:
  - Only homogeneous polynomials are (the degree is the number of atoms in the query)
  - Not every homogeneous is (e.g., \( x^2 + xy + y^2 \) is not CQ-admissible)

- In the paper we give a syntactic characterization of CQ-admissible polynomials (not that elegant, really).

- **Remark**: every polynomial is UCQ-admissible
Injective homomorphisms as a necessary condition

Using CQ-admissible polynomials it is possible to formulate an axiom $\psi$ which captures injective homomorphism as a necessary condition.
Injective homomorphisms as a necessary condition

Using CQ-admissible polynomials it is possible to formulate an axiom $\Psi$ which captures injective homomorphism as a necessary condition.

**Theorem**

The following are equivalent:

- $\mathcal{K}$ satisfies $1 + a = 1$ and $\Psi$
- $Q_1 \subseteq_{\mathcal{K}} Q_2$ iff injective homomorphism from $Q_2$ to $Q_1$
Obtain a class where containment is decidable

$\mathcal{N}[X]$

Injective Homomorphism from $Q_2$ to $Q_1$, iff $Q_1 \subseteq_K Q_2$

Class $\Psi$

$1 + a = 1$

$\mathcal{B}$
Outline

- Formalization of $\mathcal{K}$-containment
- Known results
- Some results in the paper
  - Results for homomorphisms
  - Results for homomorphic covering
  - Results for injective homomorphisms...
    and a relevant class of polynomials
  - Results for other types of mappings
Bijective homomorphisms

- What is the type of homomorphism which is sufficient condition for $\mathcal{K}$-containment for any semiring (i.e., when we relax both of the axioms)?
Bijection homomorphisms

- What is the type of homomorphism which is sufficient condition for $\mathcal{K}$-containment for any semiring (i.e., when we relax both of the axioms)?

- **Bijection homomorphism** – homomorphism with a bijection on atoms (not on variables!)
Bijective homomorphisms

- What is the type of homomorphism which is sufficient condition for $K$-containment for any semiring (i.e., when we relax both of the axioms)?

- **Bijective homomorphism** – homomorphism with a bijection on atoms (not on variables!)

- Bijective homomorphism is a sufficient and necessary for *Provenance Polynomials* (Green ’11), hence is it sufficient for any semiring.
What is the type of homomorphism which is sufficient condition for $\mathcal{K}$-containment for any semiring (i.e., when we relax both of the axioms)?

- **Bijective homomorphism** – homomorphism with a bijection on atoms (not on variables!)

- Bijective homomorphism is a sufficient and necessary for *Provenance Polynomials* (Green ’11), hence is it sufficient for any semiring.

- Using **CQ-admissible polynomials** it is possible to formulate an axiom for a class semirings for which it is a necessary condition.
Also in our paper

- Similar theorems for surjective homomorphism
- Extension to UCQs
- Complete Descriptions of CQs and UCQs
- Small model property and new procedures for semirings satisfying

\[ a + a = a \]
Future work

- Well behaved Semirings:
  \[ a + a = a \]

- Containment of CQ-admissible polynomials over various semirings

- Views over annotated databases