Graph query optimization using semi-join rewritings

Jelle Hellings\textsuperscript{1}

\texttt{jelle.hellins@uhasselt.be}

Hasselt University, Martelarenlaan 42, 3500 Hasselt, Belgium

\textsuperscript{1}Joint work with Catherine L. Pilachowski, Dirk Van Gucht, Marc Gyssens, and Yuqing Wu.
Query: ‘Great-grandparents and their friends’
Query: ‘Great-grandparents and their friends’

- (Great-grandparents, descendant): parentOf ◦ parentOf ◦ parentOf
Graphs and Graph Querying

Query: ‘Great-grandparents and their friends’

- (Great-grandparents, descendant):
  \[\text{parentOf} \circ \text{parentOf} \circ \text{parentOf}\]

- Great-grandparents:
  \[\pi_1[\text{parentOf} \circ \text{parentOf} \circ \text{parentOf}]\]
Graphs and Graph Querying

Query: ‘Great-grandparents and their friends’

- (Great-grandparents, descendant): 
  $\text{parentOf} \circ \text{parentOf} \circ \text{parentOf}$

- Great-grandparents: 
  $\pi_1[\text{parentOf} \circ \text{parentOf} \circ \text{parentOf}]$

- Complete query: 
  $\pi_1[\text{parentOf} \circ \cdots \circ \text{parentOf}] \circ \text{friendOf}$
**Graph Query Language**

| id | di | \(\ell\) | \(\ell^\perp\) | \(\pi_j[e]\) | \(\bar{\pi}_j[e]\) | \(e \circ e\) | \(e \cup e\) | \(e \cap e\) | \(e - e\) | \([e]^{\ast}\) |

- Regular Path Queries
Graph Query Language

\[
\text{id} \mid \text{di} \mid \ell \mid \ell^{-} \mid \pi_j[e] \mid \bar{\pi}_j[e] \mid e \circ e \mid e \cup e \mid e \cap e \mid e - e \mid [e]^* 
\]

- Regular Path Queries
- Nested Regular Path Queries
Graph Query Language

\[ \text{id} | \text{di} | \ell | \ell \uparrow | \pi_j[e] | \overline{\pi_j[e]} | e \circ e | e \cup e | e \cap e | e - e | [e]^* \]

- Regular Path Queries
- Nested Regular Path Queries
- FO[3] augmented with transitive closure:

  graph-navigational core of XPath, GXPath, SPARQL, ...
Query Evaluation

id | di | ℓ | ℓ^\leftrightarrow | π_j[e] | π_j[e] | e \circ e | e \cup e | e \cap e | e - e | [e]^*
Query Evaluation

\[ \text{id} \mid \text{di} \mid \ell \mid \ell^\wedge \mid \pi_j[e] \mid \overline{\pi_j[e]} \mid e \circ e \mid e \cup e \mid e \cap e \mid e - e \mid [e]^* \]

- ‘Easy to evaluate’
<table>
<thead>
<tr>
<th>id</th>
<th>di</th>
<th>ℓ</th>
<th>ℓ̃</th>
<th>π_j[e]</th>
<th>π̃_j[e]</th>
<th>e ∘ e</th>
<th>e ∪ e</th>
<th>e ∩ e</th>
<th>e − e</th>
<th>[e]^*</th>
</tr>
</thead>
</table>

- ‘Easy to evaluate’
- ‘Expensive to evaluate’
Query Evaluation

| id | di | ℓ | ℓ̅ | π_j[e] | π̅_j[e] | e ° e | e ∪ e | e ∩ e | e − e | [e]^

- ‘Easy to evaluate’
- ‘Expensive to evaluate’

Idea: add partial alternatives for ° and [·]^

\[ \pi_1[\text{parentOf} ° \text{parentOf} ° \text{parentOf}] ° \text{friendOf} \]

can be rewritten into

\[ \pi_1[\text{parentOf} ⋉ (\text{parentOf} ⋉ \text{parentOf})] ⋉ \text{friendOf}. \]
Query Optimization by rewriting?

Problem

parentOf ◦ parentOf ◦ parentOf

is not equivalent to

parentOf ∨ (parentOf ∨ parentOf).

[Diagram showing relationships between Alice, Bob, Eve, Carol, Dan, and Faythe, with arrows indicating parentOf and friendOf relationships.]
Query Optimization by rewriting?

Problem

parentOf \circ parentOf \circ parentOf

is not equivalent to

parentOf \ltimes (parentOf \ltimes parentOf).

Solution

\textit{j-test-equivalent rewriting}: we have \( e_1 \equiv_j e_2 \), if, for every graph \( G \),

\[ \pi_j[e_1](G) = \pi_j[e_2](G). \]
Rewrite composition and transitive closure

- Rewrite $\circ$ into $\times$ and $\otimes$
- Rewrite $[\cdot]^*$ into $\text{fp}_{j,\mathcal{N}}[\cdot; \cdot]$ (fixpoint iteration)

$$\text{id} \mid \text{di} \mid \ell \mid \ell^\perp \mid \pi_j[e] \mid \overline{\pi}_j[e] \mid e \circ e \mid e \cup e \mid e \cap e \mid e - e \mid [e]^* \mid$$

$$e \otimes e \mid e \otimes e \mid \mathcal{N} \mid \text{fp}_{j,\mathcal{N}}[e; e]$$
Rewrite composition and transitive closure

- Rewrite \( \circ \) into \( \times \) and \( \ltimes \)
- Rewrite \([\cdot]^*\) into \( \text{fp}_{j,N}[\cdot; \cdot]\) (fixpoint iteration)

\[
\begin{align*}
\text{id} & \mid \text{di} \mid \mathcal{L} \mid \mathcal{L}^- \mid \pi_j[e] \mid \overline{\pi}_j[e] \mid e \circ e \mid e \cup e \mid e \cap e \mid e - e \mid [e]^* \mid \\
e \times e & \mid e \times e \mid N \mid \text{fp}_{j,N}[e; e]
\end{align*}
\]

Analysis

- FO[2]
Rewrite composition and transitive closure

- Rewrite $\circ$ into $\ltimes$ and $\ltimes$
- Rewrite $[\cdot]^*$ into $\text{fp}_{j,\pi}[\cdot; \cdot]$ (fixpoint iteration)

\[
\text{id} | \text{di} | \ell | \ell^\wedge | \pi_j[e] | \pi_j[e] | e \circ e | e \cup e | e \cap e | e - e | [e]^* | e \ltimes e | e \ltimes e | \pi | \text{fp}_{j,\pi}[e; e]
\]

Analysis

- FO[2] and FO[2]-like recursion
- For $j$-test-equivalent rewriting: only restrictions on $\cap$ and $-$
Rewrite composition and transitive closure

- Rewrite $\circ$ into $\times$ and $\boxtimes$
- Rewrite $[\cdot]^*$ into $\text{fp}_{j,\mathcal{M}}[\cdot; \cdot]$ (fixpoint iteration)

\[
\begin{align*}
  \text{id} & \mid \text{di} & \ell & \ell^\perp & \pi_j[e] & \pi_j[e] & e \circ e & e \cup e & e \cap e & e - e & [e]^* \\
  e \times e & \mid e \times e & \mathcal{M} & \text{fp}_{j,\mathcal{M}}[e; e]
\end{align*}
\]

Analysis

- FO[2] and FO[2]-like recursion
- For $j$-test-equivalent rewriting: only restrictions on $\cap$ and $-$
- Rewriting is sound and ‘complete’
Rewrite composition and transitive closure

- Rewrite $\circ$ into $\times$ and $\ltimes$
- Rewrite $[\cdot]^{*}$ into $\text{fp}_{j,\mathfrak{M}}[\cdot; \cdot]$ (fixpoint iteration)

$$
\begin{align*}
\text{id} & \mid \text{di} \mid \ell \mid \ell^{-} \mid \pi_{j}[e] \mid \pi_{j'}[e] \mid e \circ e \mid e \cup e \mid e \cap e \mid e - e \mid [e]^{*} \mid e \ltimes e \mid e \ltimes e \mid \mathfrak{M} \mid \text{fp}_{j,\mathfrak{M}}[e; e]
\end{align*}
$$

Analysis

- FO[2] and FO[2]-like recursion
- For $j$-test-equivalent rewriting: only restrictions on $\cap$ and $-$
- Rewriting is sound and ‘complete’
- Rewriting results in a ‘small’ query: number of steps needed to evaluate the result is twice the length of the original query
Future Work

- Study (small extensions of) FO$[2]$ in more detail
- Further query optimization using information on the data
- Apply similar techniques to relational databases (SQL)
The transitive closure query

\[ \pi_1[[\text{parentOf} \circ \pi_1[\text{researcherAt}]]^* \circ \text{ownsPet}] \]
The transitive closure query

\[ \pi_1[[\text{parentOf} \circ \pi_1[\text{researcherAt}]^* \circ \text{ownsPet}]] \]

is equivalent to the FO[2]-like query

\[ \text{fp}_{1,\pi_1}[\text{parentOf} \sqcup \pi_1[\text{researcherAt}] \sqcup \pi; \text{ownsPet}]. \]