Massively Multi-Query Join Processing in Publish/Subscribe Systems (SIGMOD’07)

Ying Zheng
Instructor: Jun Yang
2008-03-18

Outline

• Motivation
• XSCL Query Language
• Two-Stage Query Processing
• Experiment
• Conclusion

Motivation

• Inter-document Queries in pub/sub systems
  – Different XML document joins based on values in their nodes, including attributes and text
  – E.g. Return a book announcement, followed by a blog article from one of its authors with the same title as the book.

XSCL Query Language

• SELECT, FROM, PUBLISH
• XPath operator
• Variable binding construct
• Join operator (conjunctive predicates, window size)
  – JOIN
  – FOLLOWED BY
• Example
  
  $/\text{book}://\text{author}://\text{title}$
  
  FOLLOWED BY $(x2=x5 \text{ AND } x3=x6, T1)$
  $/\text{blog}://\text{author}://\text{title}$(T2)

Two-Stage Query Processing

Query Example:

$\text{book}://\text{author}://\text{title}$

FOLLOWED BY $(x2=x5 \text{ AND } x3=x6, T1)$
$\text{blog}://\text{author}://\text{title}$(T2)

XPath Evaluation: YFilter
**Xpath Evaluation: Generate DB**

**Query:**
```
Table: Rbin
FOLLOWED BY(x=2 AND x<5, T1)
```

**Table: Rbin**
```
<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>z</th>
<th>w</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
```

**Query Template**
```
Q1: FOLLOWED BY(x=2 AND x<5, T1)
Q2: FOLLOWED BY(x=2 AND x<5, T1)
```

**Join Processing & Optimization**

**Query:**
```
Table: Rbin
FOLLOWED BY(x=2 AND x<5, T1)
```

**Table: Rbin**
```
<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>z</th>
<th>w</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
```

**Experiment**

**Conclusion**

- Inter-document Queries in pub/sub systems
- XSCL Query Language
  - XML structure
  - Equivalence predicates for values on the leaves
- Two-stage Query Processing
  - XPath Evaluation
  - Join Processing
  - Optimization: View Materialization