QURSED:
Querying and Reporting
Semistructured Data

Yannis Papakonstantinou
Michalis Petropoulos
University of California, San Diego

Vasilis Vassalos
New York University

June 2002
Overview

- Query Forms and Reports
  - Challenges of Semistructured Data
- The QURSED system
  - Architecture
- Technical foundation
  - Tree Query Language (TQL)
  - Query Set Specification (QSS)
- QURSED Editor
Exporting DBMSs on the Web

- XML views and schemas
- XQuery behind the scenes
- Need for web-based interfaces
Web and Databases Effort

- Data intensive Web site generators
  - Strudel
    - Forms as functions on edges/links
  - Araneus
  - Autoweb

- Declarative
- Separation of content, structure and presentation
Query Forms and Reports

Requirements

• Handle semistructureness
  – Powerful query forms and reports
• Be declarative
  – Separate logic from presentation
• Encode compactly a large number of queries
  – Compared to a set of query templates
• Visual interface for the developer
  – Programming should NOT be a requirement
Query Forms and Reports
Query Forms and Reports

QURSED Approach

- XML Schema-driven
- Declarative!
  - Separation of content & presentation

- Editor
  - Visual actions to declarative specifications
  - Automatic construction of report pages

- Query Set Specification (QSS)
  - Large set of parameterized queries
  - Compact representation

- Engine
  - Automatic query formulation
  - Direct result construction
Query Forms and Reports

QURSED Editor

XML Schema

Query Forms and Reports
Developing Query Forms from the XML Schema
Developing Reports from the XML Schema
Tree Query Language (TQL)

- Condition tree
- Schema-driven generation

- Result tree
Tree Query Language (TQL)

AND

PROT1 = “NEMA3” OR
PROT1 = “NEMA1”

sensors

manufacturer

name

product

specs

sensing_distance

body_type

diameter

barrel_style

protection_ratings

protection_rating

$S$

$NAME$

$PROD$

manufacturer

name

“Turck”

product

part_number

“A123”

specs

sensing_distance

“11”

body_type

cylindrical

diameter

“17”

barrel_style

“Smooth”

protection_ratings

protection_rating

“NEMA1”

protection_rating

“NEMA3”

$PROT1$

$PROT1$

sensors

manufacturer

name

“Turck”

product

part_number

“B123”

specs

sensing_distance

“25”

body_type

rectangular

height

“10”

width

“30”

protection_ratings

protection_rating

“NEMA3”

protection_rating

“NEMA4”

<table>
<thead>
<tr>
<th>$NAME$</th>
<th>$PROD$</th>
<th>$PART$</th>
<th>$PROT1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turck</td>
<td>product</td>
<td>A123</td>
<td>NEMA1</td>
</tr>
<tr>
<td>Turck</td>
<td>product</td>
<td>A123</td>
<td>NEMA3</td>
</tr>
<tr>
<td>Turck</td>
<td>product</td>
<td>B123</td>
<td>NEMA3</td>
</tr>
</tbody>
</table>
Tree Query Language (TQL)
TQL Semantics

Condition Tree

- Conjunctive Condition Trees
  - **OR-Removal** Algorithm
  - Transformation Rules
TQL Semantics

Result Tree
Tree Query Language (TQL)

- Translated to XQuery
  - By QURSED Run-Time Engine
  - **TQL2XQuery Algorithm**
  - Syntax directed translation
  - Tree patterns in TQL to nested FOR-WHERE-RETURN expressions in XQuery
Query Set Specification (QSS)

Condition Tree Generator
- Parameterized boolean expressions
- Multiple boolean expressions per AND node
- Condition fragments

$sensing_distance \leq $#DIST$
$sensing_distance \leq $#DIMX$ AND $sensing_distance \leq $#DIMY$
$protection_rating \leq $#PROT1$
$height \leq $#HEI$ AND $width \leq $#WID$
$barrel_style \leq $#BAR$
$rectangle \leq $#REC$
$cylindrical \leq $#CYL$
$dimension \leq $#DIA$
$protection_rating \leq $#PROT1$
$manufacturer \leq $#S$
$name \leq $#NAME$
$product \leq $#PROD$
$sensors \leq $#S$
$specs \leq $#S$
$body_type \leq $#S$
$OR$
$AND$
$AND$
$AND$

$f_1 f_2 f_3$
Dependencies

$BODY = \#BODY$

$\text{AND}$

$\text{AND}$

$\text{AND}$

$sensors$

$\text{body\_type}$

$\text{cylindrical}$

$diameter \rightarrow \#\text{DIA}$

$\text{barrel\_style} \rightarrow \#\text{BAR}$

$\text{rectangular}$

$\text{height} \rightarrow \#\text{HEI}$

$\text{width} \rightarrow \#\text{WID}$

$f_1$

$f_2$

$f_3$

$\text{<f2, } \#\text{BODY} = \text{"Cylindrical", } \{f_1\}>$

$\text{Mechanical}$

Body Type

Cylindrical

Diameter

mm

Barrel Style

No preference

$\text{<f3, } \#\text{BODY} = \text{"Rectangular", } \{f_1\}>$
Dependencies

\[ f_1 \]
\[ f_2 \]
\[ f_3 \]

\[ <f_2, \#BODY = "cylindrical", \{f_1\}> \]
\[ <f_3, \#BODY = "rectangular", \{f_1\}> \]

- Dependencies Graph
- Resolution algorithm based on topological sort
Run-time: QSS to TQL Queries

Query:

- $\text{NAME} = \#\text{NAME}$
- $\text{DIST} \leq \#\text{DIST}$
- $\text{NAME} = \text{Turck}$
- $\text{DIST} \leq 6$
- $\text{NAME} = \text{Turck}$
Run-time: QSS to TQL Queries

- Constants from Query Form Page
- XHTML Report Page

QURSED Run Time Engine

- Instantiate parameters in CTG with constants
- Find Active Condition Fragments
- Resolve Dependencies
- Union of Active Condition Fragments to TQL Query
- TQL Query to XQuery Expression

XQuery Expressions

XQuery Engine

XML/XHTML
QURSED Editor

Building Query/Visual Association
From visual actions to QSS

Choice of schema element $e$ means
- Addition of $e$ to the $CTG$
- Addition of the $e$ path to the $CTG$
- Creation of a name variable for $e$

sensors/manufacturer/*
= man_name_select
Disjunction

\[(\$DIA \leq \#DIMX \text{ AND } \$DIA \leq \#DIMY) \text{ OR } (\$HEI \leq \#DIMX \text{ AND } \$WID \leq \#DIMY)\]
• Creation of disjunctive condition triggers transformation of the Condition Tree Generator
  – ORNodes Algorithm
QURSED Editor

Building Reports

Elements to Appear on Report

Group By Mapping

Element Mapping
QURSED Editor

Result Tree

AND
sensors
manufacturer
name
$NAME
product
$PROD
part_number
$PART
specs
sensing_distance
$DIST

$NAME

GROUPBY ($NAME)

MANUFACTURER

GROUPBY ($MAN)

PRODUCT

GROUPBY ($PROD)

cylindrical

GROUPBY ($CYL)

RECTANGULAR

GROUPBY ($REC)

html

GROUPBY ($BAR)

GROUPBY ($HEI)

GROUPBY ($WID)

GROUPBY ($DIA)
More Features

- Expandable schema
  - Multiple copies/variables for repeatable elements
- Optional elements
- Sort-by options
- Template-driven construction of report pages
  - Element mappings
  - Group-by mappings
- Detailed list of visual actions of QURSEDE Editor
QURSED Contributions

- The first web-based generator of powerful query forms and reports for semistructured XML data
- Declarative
  - Separates querying functionality and presentation
- Handles semistructureness
  - Disjunction
- Technical foundation
  - XML Schema, QSS, TQL, XQuery
- QURSED Editor
  - Visual actions “translated” to QSS and query/visual association
  - Automates report construction for heterogeneous data
Related work

• Web-based Form and Report Generators
  – Macromedia Ultradev, Coldfusion, Microsoft Visual InterDev
  – Excellent for flat uniform relational tables
  – Visual query formulation paradigm allows the specification of projections, sort-bys, simple conditions
  – However, the development of form and report pages for semistructured data requires substantial programming effort

• Visual Querying Interfaces
  – EquiX, BBQ, VQBD, Lorel’s DataGuide-driven GUI, PESTO
  – Excellent visual paradigm for the formulation of fairly complex queries
  – The goal is the development of a query or a query template
  – User needs to be familiar with database models and schemas
Questions and Answers