

QURSED: Querying and Reporting Semistructured Data

Yannis Papakonstantinou
Michalis Petropoulos

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Vasilis Vassalos

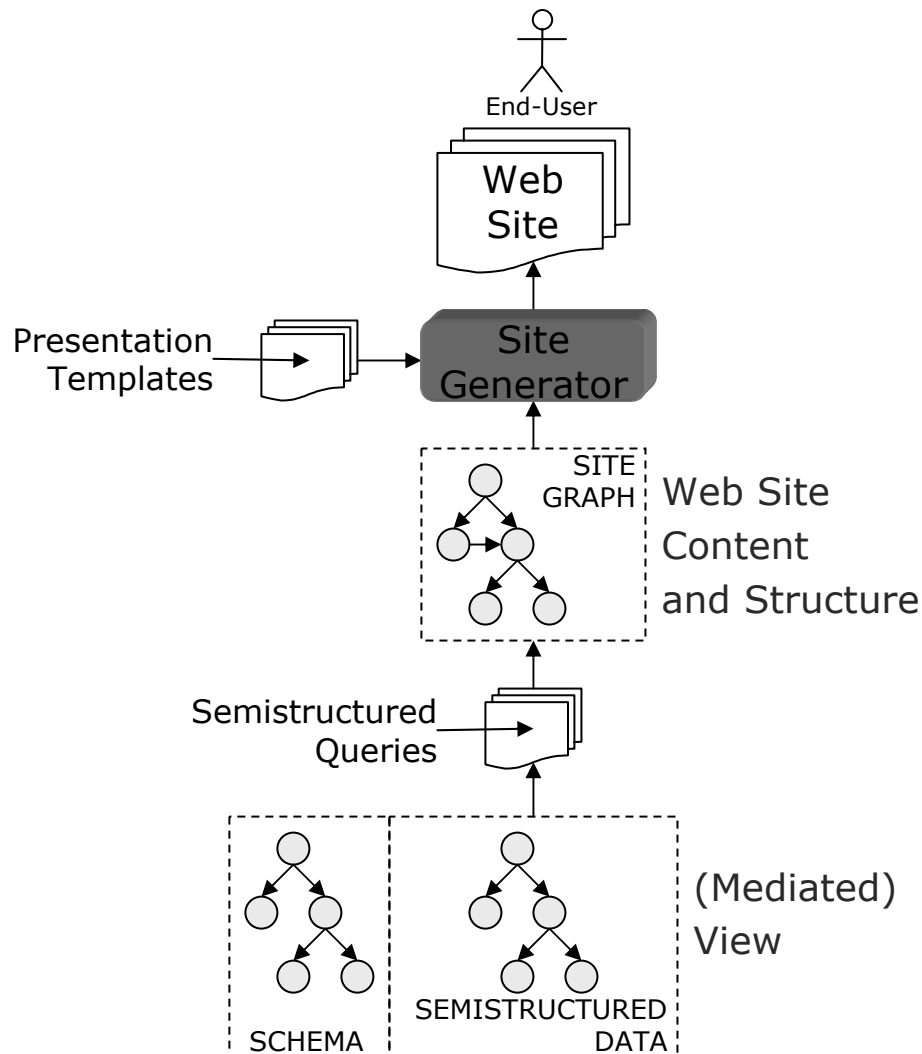
NEW YORK UNIVERSITY

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

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 Form and Report Pages - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Sensors

General

Manufacturer:

Balluff

Baumer

Turck

Sensing Distance: mm

Protection Rating 1:

Protection Rating 2:

Operating Temperature: to °C

Mechanical

Body Type:

Dimension X: mm






Dimension Y: mm

Results

Results/page:

Sort By Options: Sensing Distance

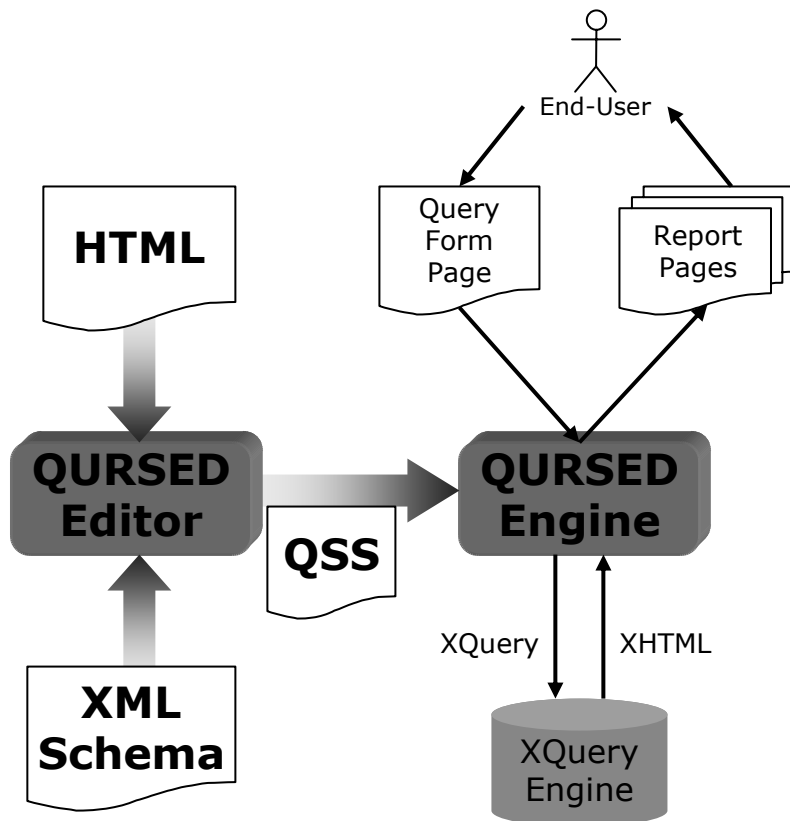
DESC-Manufacturer

Next 10				Previous 10	
Image	Manufacturer	Part Number	Protection Ratings	Sensing Distance mm	Body Type
	<input type="text" value="All"/>		<input type="text" value="All"/>	<input type="text" value="All"/>	
	<input type="text" value="All"/> Balluff Turck	BC 3-M12-AN6X	NEMA1 NEMA3 NEMA4	6.0	Cylindrical Diameter mm Barrel Style 15 Smooth
	Turck	BC 3-M12-AP6X	NEMA3	6.0	Cylindrical Diameter mm Barrel Style 19 Smooth
	Turck	BC 5-Q08-AN6X2	NEMA3 NEMA4	7.0	Rectangular Height mm Width mm 14 9
	Turck	BC 5-Q08-AP6X2	NEMA3 NEMA6 NEMA11	7.5	Rectangular Height mm Width mm 10 35
	Turck	BC 5-S18-AN4X		10.0	Rectangular Height mm Width mm 15 10
	Turck	BC 5-S18-AP4X	NEMA1 NEMA3	10.6	Rectangular Height mm Width mm

My Computer

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

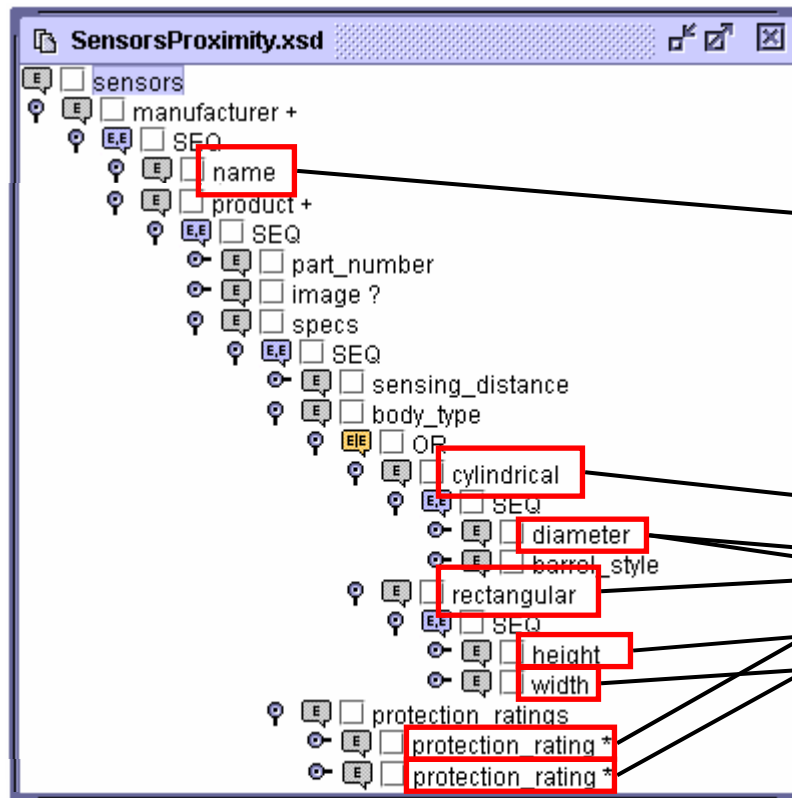
Query Forms and Reports

QURSED Editor

The screenshot displays the QURSED Editor interface, which is used for creating query forms and reports. The interface is divided into several panels:

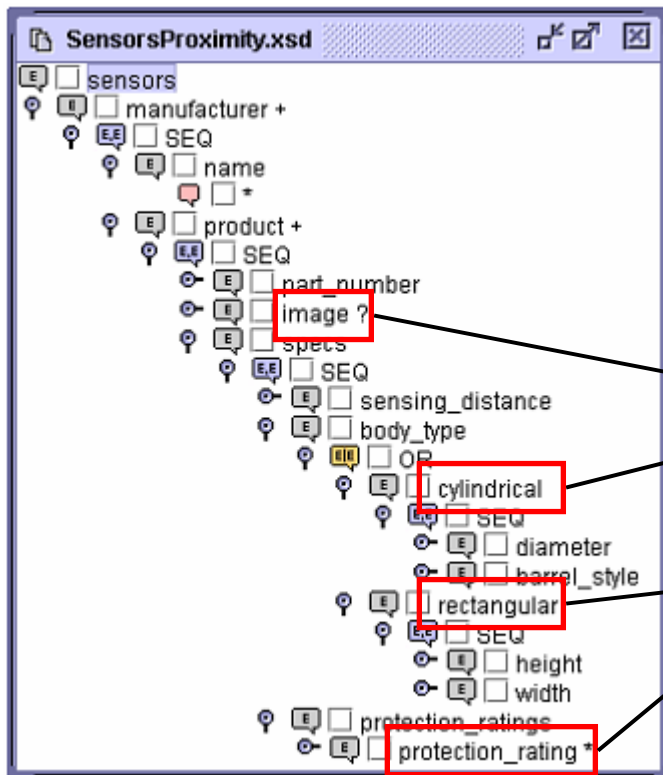
- XML Schema:** Located on the left, it shows a tree view of the XML schema for `SensorsProximity.xsd`. The root element is `sensors`, which contains several child elements, including `manufacturer`, `product`, `specs`, `sensing_distance`, `body_type`, `protection_ratings`, and `operating_temp`. A blue arrow points to the `manufacturer` element.
- Query Form:** Located in the center, it shows a table of condition fragments. The table has two columns: `ID` and `Condition Fragment ID`. The first row is selected, showing `manufacturer_name` as the condition fragment and `protection_rating_1` as the ID. Below the table is an **Expression Editor** window with tabs for `Arithmetic`, `Comparison`, `Boolean`, `Constant`, and `Custom`. The `Comparison` tab is active, and the expression `sensors/manufacturer/name/* = man_name_select` is entered in the text field. The `Arg OP Arg` checkbox is checked. `OK` and `Cancel` buttons are at the bottom.
- Query Forms and Reports:** Located on the right, it shows a tree view of the query form and report structure. The root element is `html`, which contains a `body` element. The `body` element contains a `form` element, which contains a `select` element. The `select` element contains several options, including `man_name_select`, `Baumer`, `Turck`, `prot_rating_1_select`, `NEMA3`, `prot_rating_2_select`, and `NEMA1`. A blue arrow points to the `man_name_select` option.

Developing Query Forms from the XML Schema



The screenshot shows the 'Query Form and Report Pages' window. The form is titled 'Sensors' and has a 'General' section with the following fields: 'Manufacturer' (a dropdown menu with 'Turck' selected), 'Sensing Distance' (a text input with '6' and a unit dropdown with 'mm'), 'Protection Rating' (a dropdown menu with 'No preference' selected), and 'Protection Rating' (another dropdown menu with 'No preference' selected). The 'Mechanical' section has 'Body Type' (a dropdown menu with 'Rectangular' selected), 'Height' (a text input with a unit dropdown with 'mm'), and 'Width' (a text input with a unit dropdown with 'mm'). The 'Results' section has 'Results/page' (a dropdown menu with '10' selected), 'Sort By Options' (a dropdown menu with 'Sensing Distance' selected and 'DESC' selected), and 'DESC-Manufacturer' (a dropdown menu with 'up' and 'dn' buttons). There are 'Reset' and 'Execute' buttons at the bottom of the form.






Developing Reports from the XML Schema



C:\Documents and Settings\Administrator.ALEXANDROS\Desktop\example\Results3.htm - Micr...

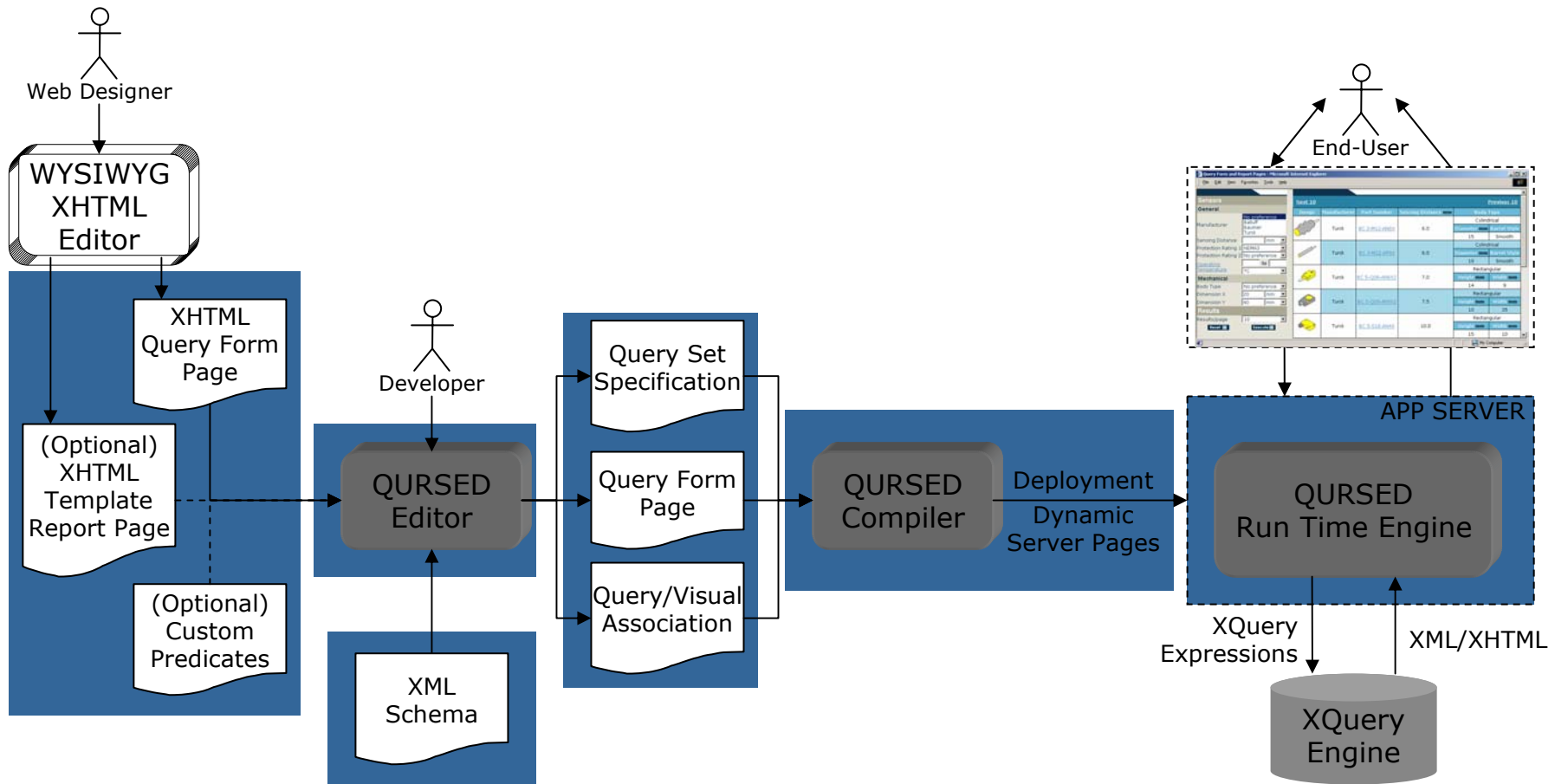
File Edit View Favorites Tools Help

Next 10 Previous 10

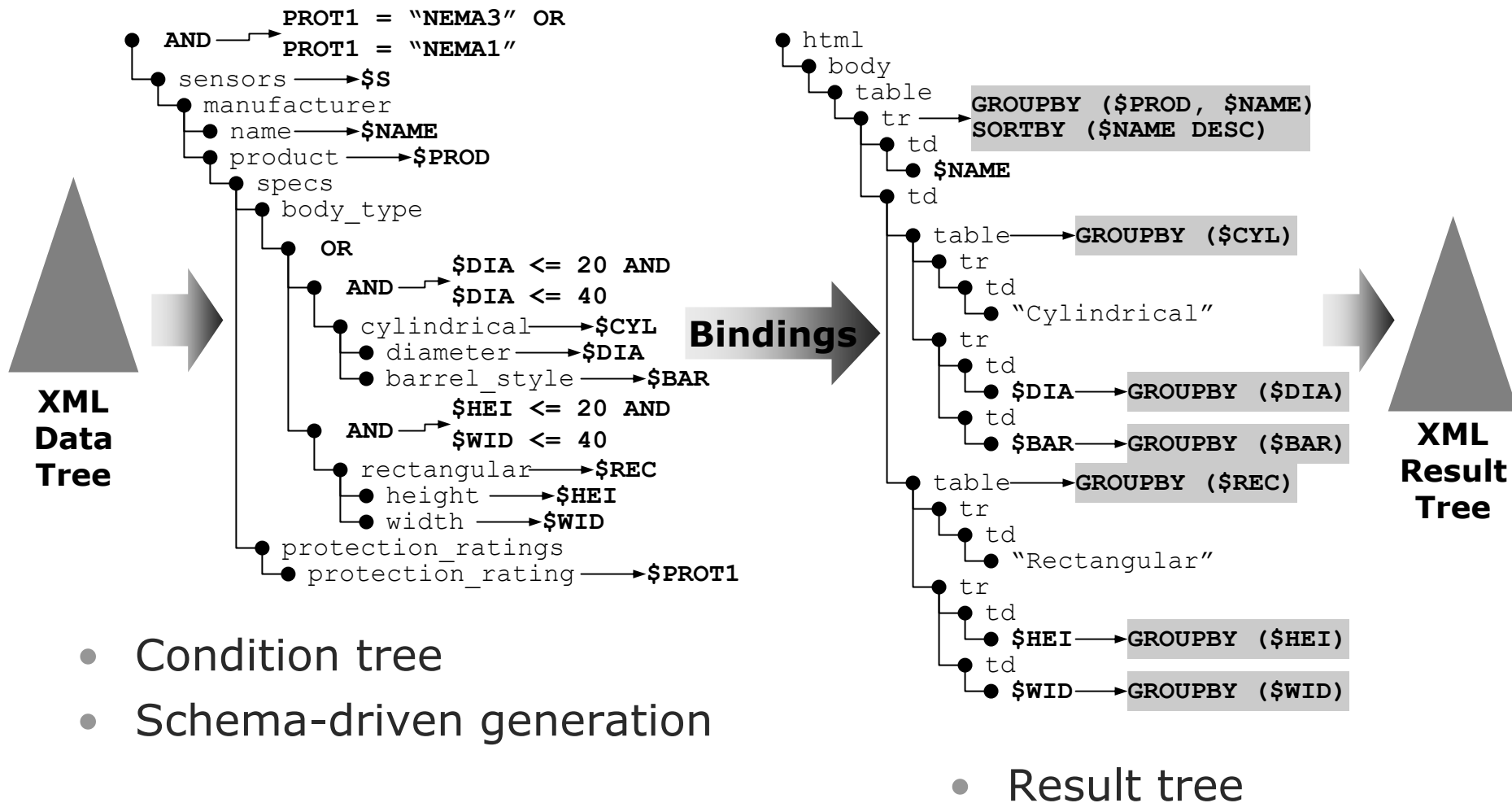
Image	Manufacturer	Part Number	Protection Ratings	Sensing Distance mm	Body Type
	Turck	BC 3-M12-AN6X	NEMA1 NEMA3 NEMA4	6.0	Cylindrical Diameter mm Barrel Style 15 Smooth
	Turck	BC 3-M12-AP6X	NEMA3	6.0	Cylindrical Diameter mm Barrel Style 19 Smooth
	Turck	BC 5-Q08-AN6X2	NEMA3 NEMA4	7.0	Rectangular Height mm Width mm 14 9
	Turck	BC 5-Q08-AP6X2	NEMA3 NEMA6 NEMA11	7.5	Rectangular Height mm Width mm 10 35
	Turck	BC 5-S18-AN4X		10.0	Rectangular Height mm Width mm 15 10

My Computer

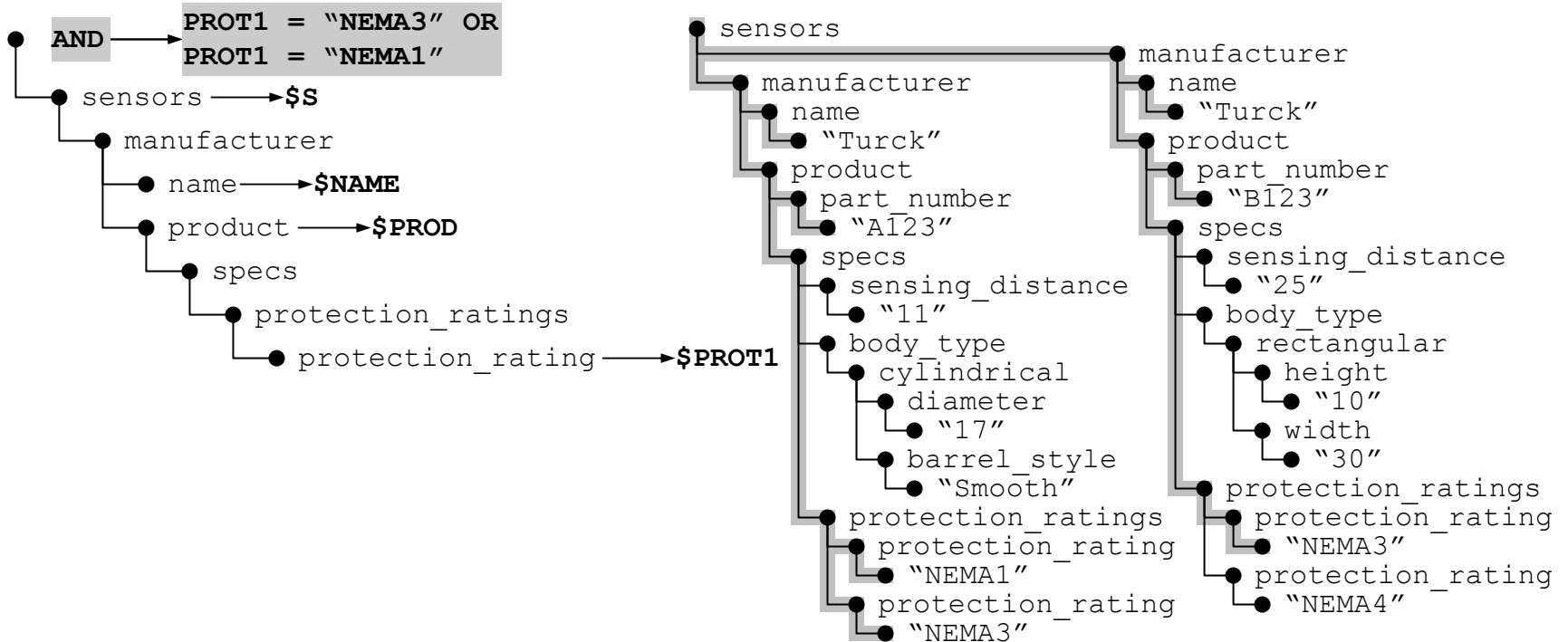
QURSED System Architecture



Tree Query Language (TQL)



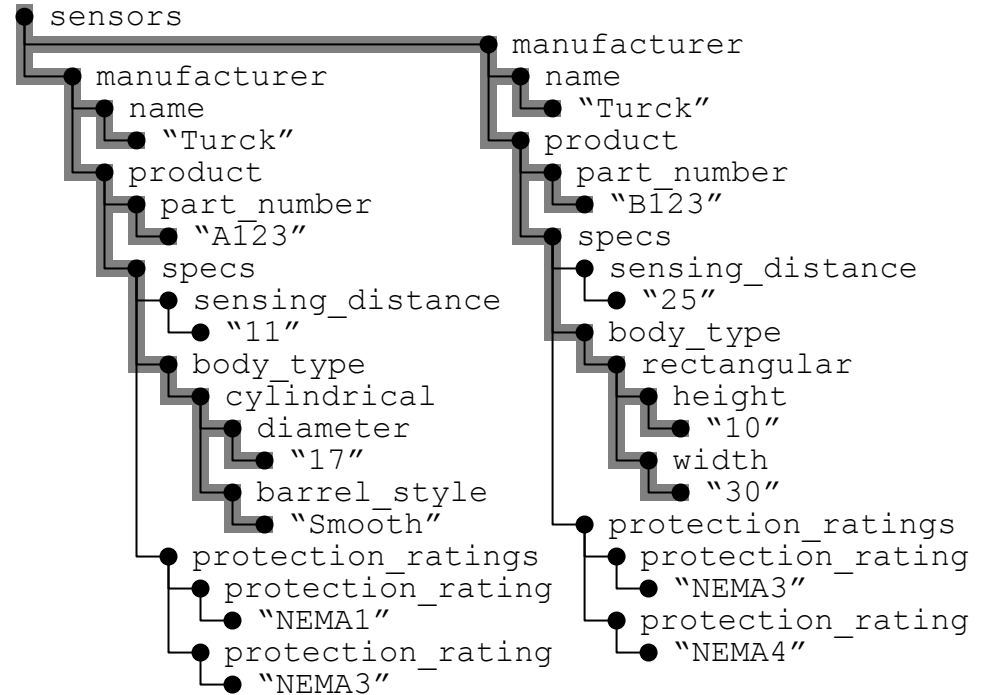
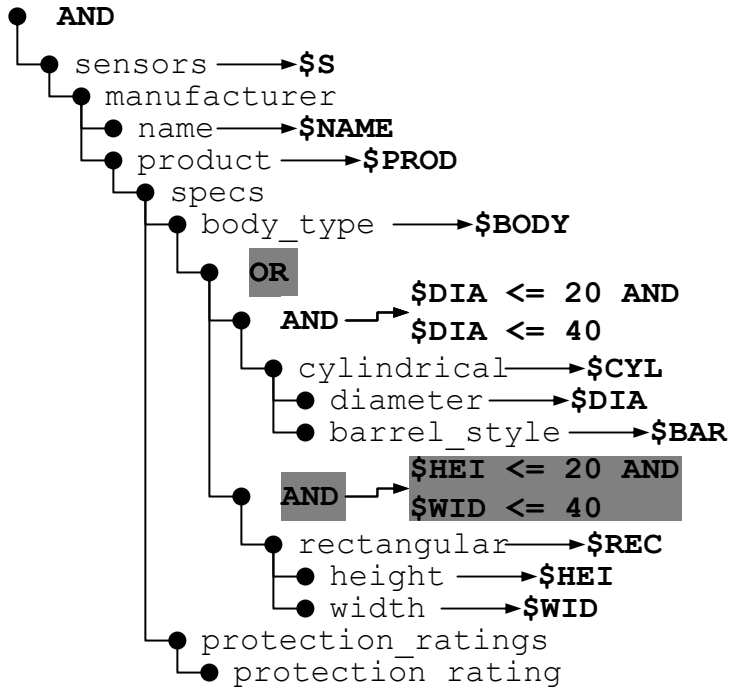
Tree Query Language (TQL)



\$NAME	\$PROD	\$PART	\$PROT1
Turck	<ul style="list-style-type: none"> ● product └─● part_number └─● "A123" 	A123	NEMA1
Turck	<ul style="list-style-type: none"> ● product └─● part_number └─● "A123" 	A123	NEMA3
Turck	<ul style="list-style-type: none"> ● product └─● part_number └─● "B123" 	B123	NEMA3



Tree Query Language (TQL)



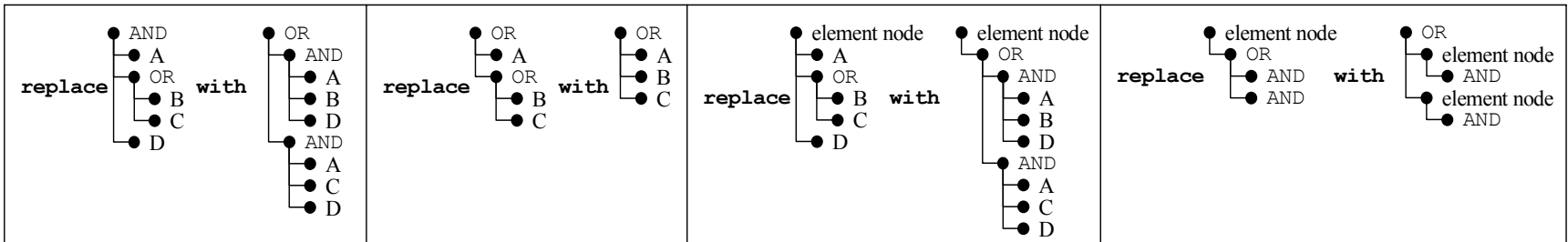
\$NAME	\$PROD	\$PART	\$BODY	\$CYL	\$DIA	\$BAR
Turck	<ul style="list-style-type: none"> ● product └─● part_number └─● "A123" 	A123	cylindrical	<ul style="list-style-type: none"> ● cylindrical └─● diameter └─● "17" 	17	Smooth

\$NAME	\$PROD	\$PART	\$BODY	\$REC	\$HEI	\$WID
Turck	<ul style="list-style-type: none"> ● product └─● part_number └─● "B123" 	B123	rectangular	<ul style="list-style-type: none"> ● rectangular └─● height └─● "10" 	10	30

TQL Semantics

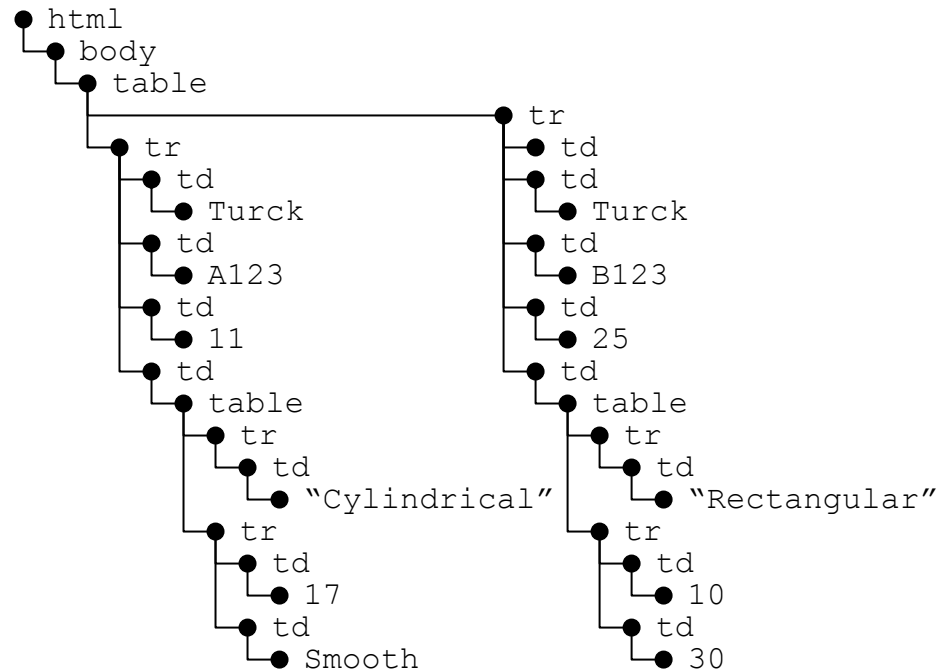
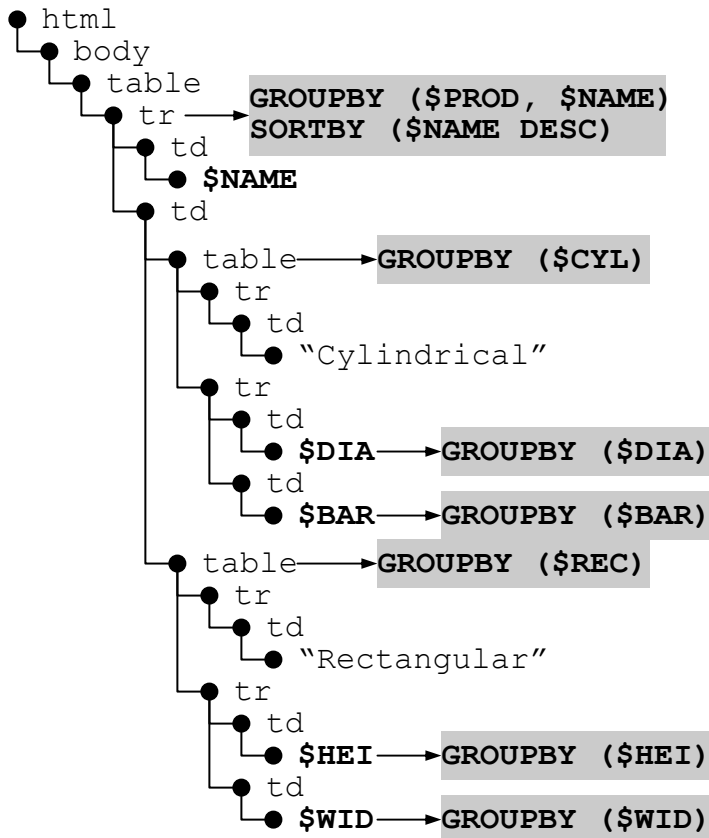
Condition Tree


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



TQL Semantics

Result Tree

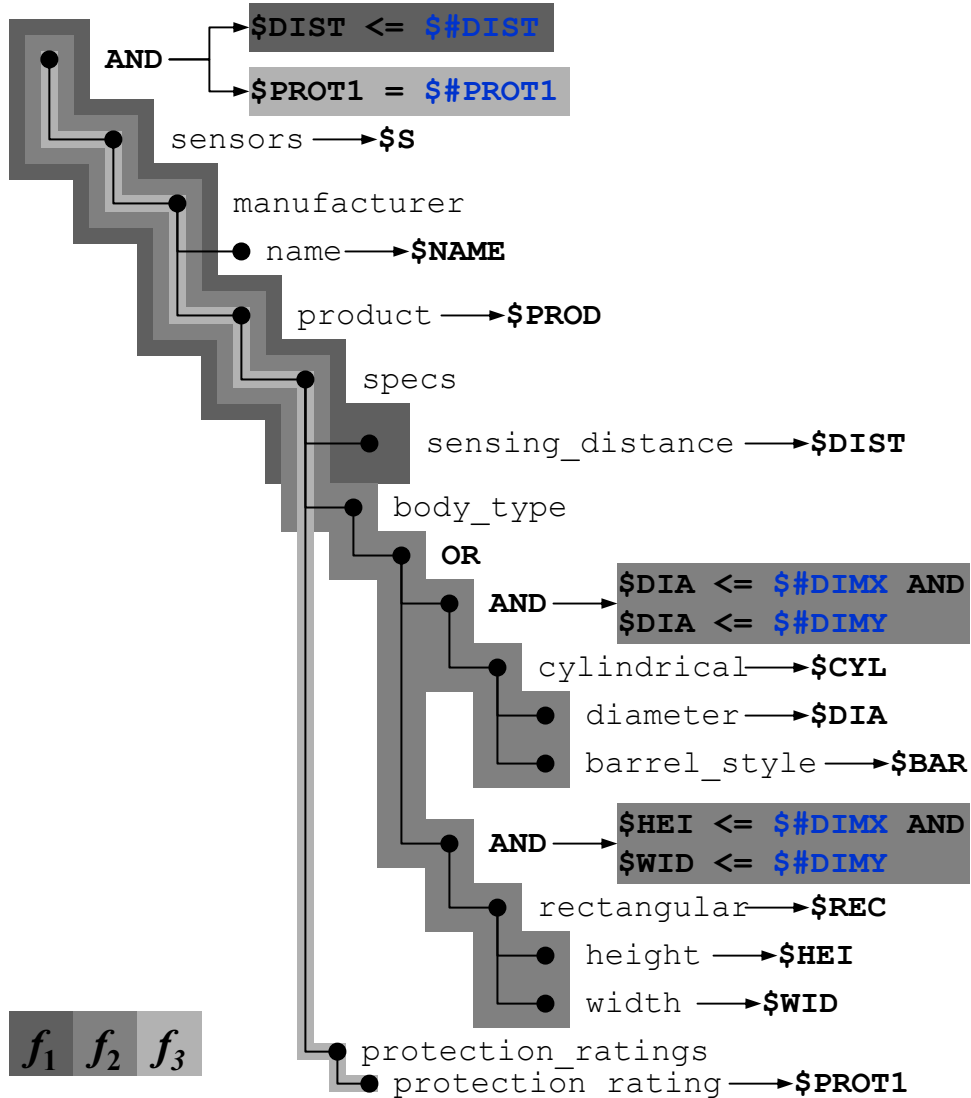


Name	Part Number	Image	Sensing Distance	Cylindrical	
Turck	A123		11.0	Diameter mm	Barrel Style
				17	Smooth
Turck	B123		25.0	Rectangular	
				Height mm	Width mm
				10	30

Tree Query Language (TQL)

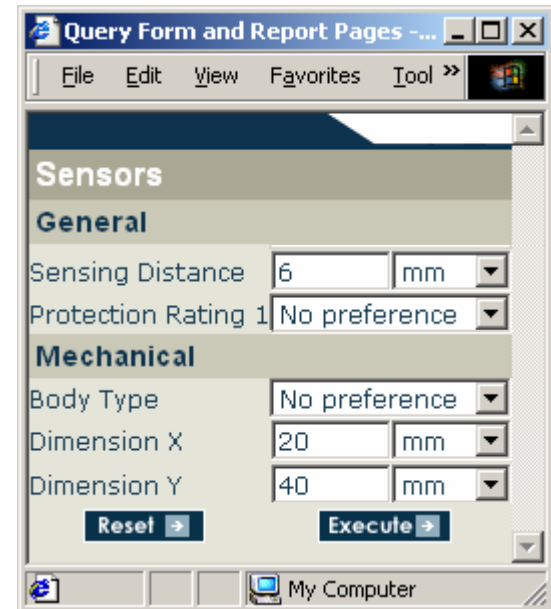
- Translated to XQuery
 - By QURSED Run-Time Engine
 - **TQL2XQuery Algorithm**
 - Syntax directed translation
 - Tree patterns in TQL to nested FOR-WHERE-RETURN expressions in XQuery

Query Set Specification (QSS)

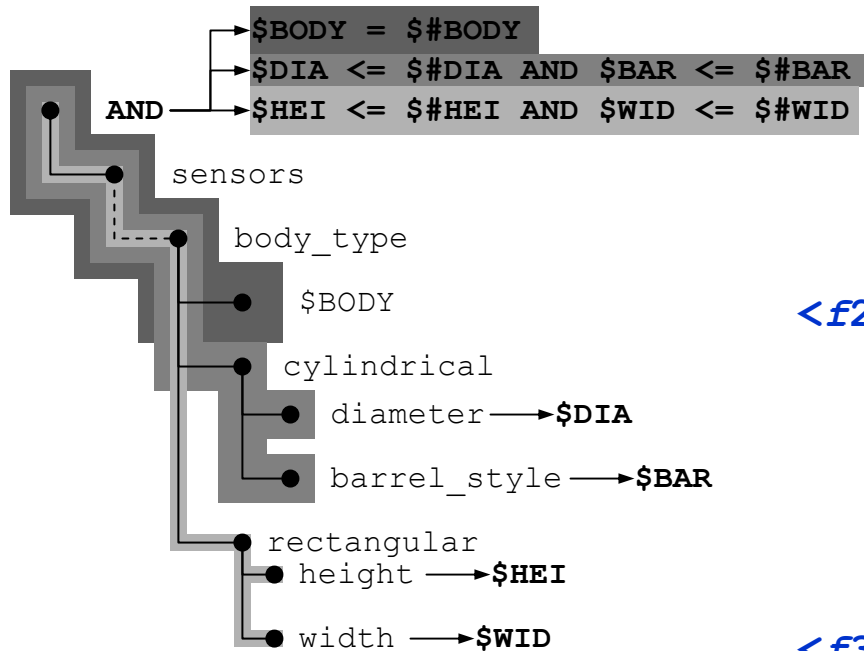


Condition Tree Generator

- Parameterized boolean expressions
- Multiple boolean expressions per AND node
- **Condition fragments**



Dependencies



Mechanical	
Body Type	Cylindrical
Diameter	mm
Barrel Style	No preference

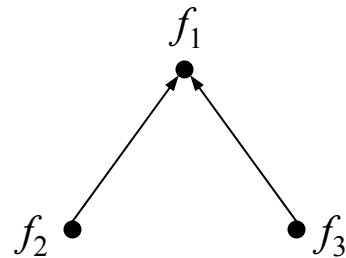
`<f2, $#BODY = "Cylindrical", {f1}>`

Mechanical	
Body Type	Rectangular
Height	mm
Width	mm

`<f3, $#BODY = "Rectangular", {f1}>`

`f1 f2 f3`

Dependencies

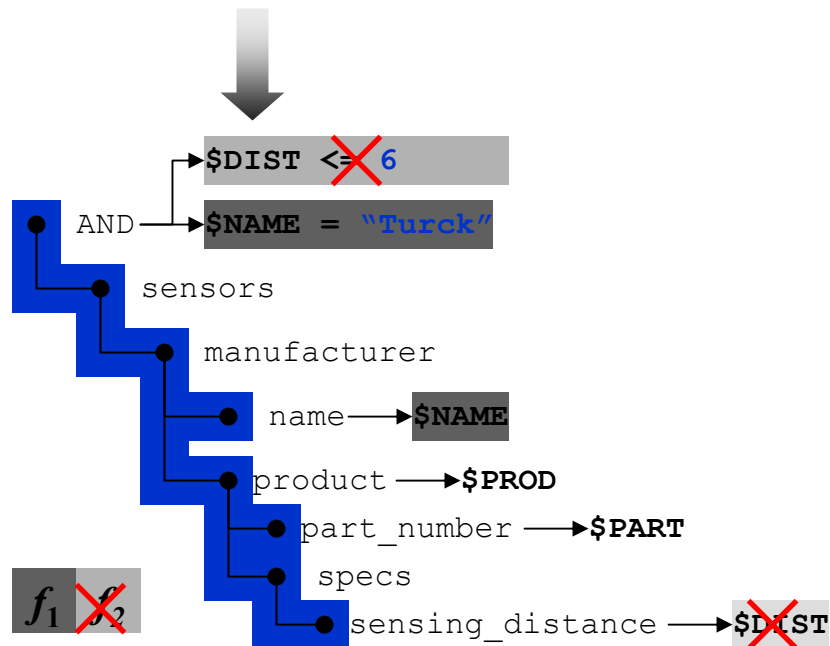
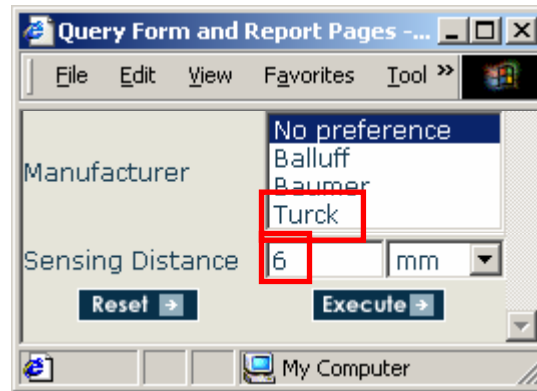


`<f2, $#BODY = "cylindrical", {f1}>`

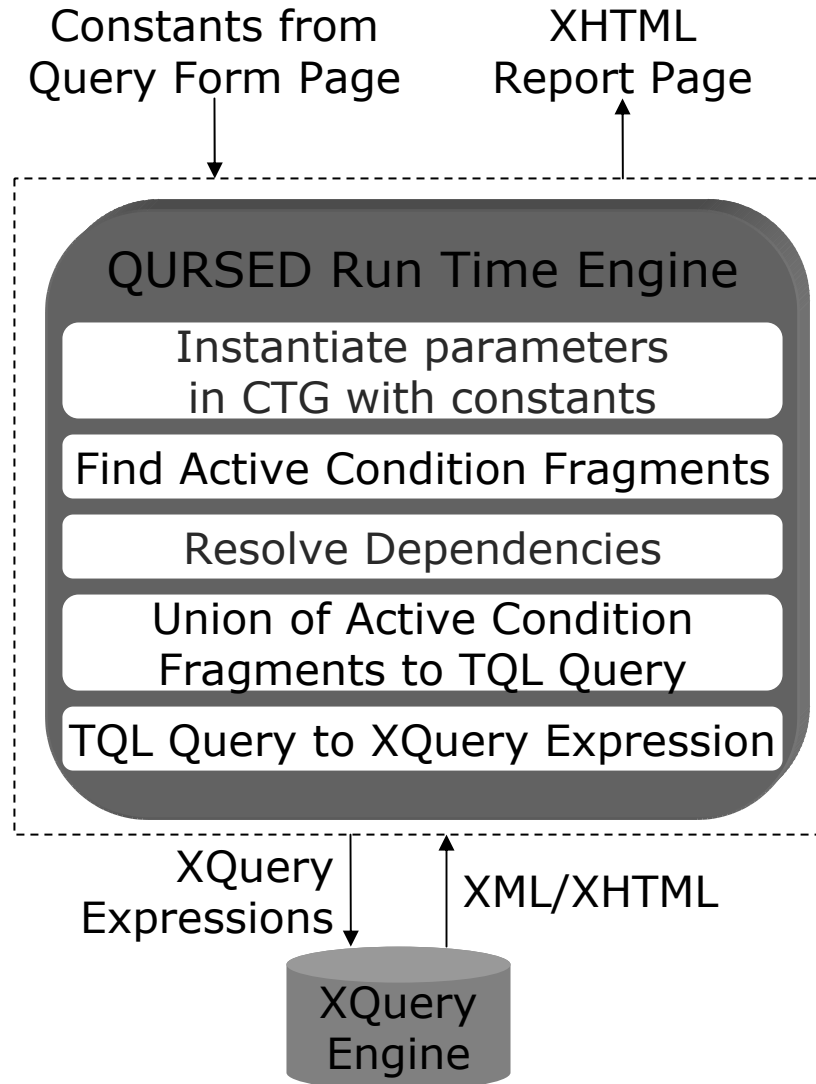
`<f3, $#BODY = "rectangular", {f1}>`

- Dependencies Graph
- Resolution algorithm based on topological sort

Run-time: QSS to TQL Queries



Run-time: QSS to TQL Queries



QURSED Editor

Building Query/Visual Association

Condition Fragment List

Data Path

Form Control

Predicate

Editor

File View Action Deploy Help

New Open Save Build Report Deploy

Data Source(s)

SensorsProximity.xsd

sensors

manufacturer +

SEQ

name

product +

SEQ

part_number

image ?

specs

SEQ

sensing_distance

body_type

protection_ratings

protection_rating *

protection_rating *

operating_temp

Query Form Report

Condition Fragments

ID

manufacturer_name

protection_rating_1

protection_rating_2

Dependencies

ID Condition Fragment ID

Expression Editor

Arithmetic Comparison Boolean Constant Custom

- Arg OP Arg

sensors/manufacturer/name* = man_name_select

OK Cancel

Query Form Page Template Report Page

html

body

form

select

name

man_name_select

option

option

Baumer

option

Turck

prot_rating_1_select

option

option

NEMA3

option

select

name

prot_rating_2_select

option

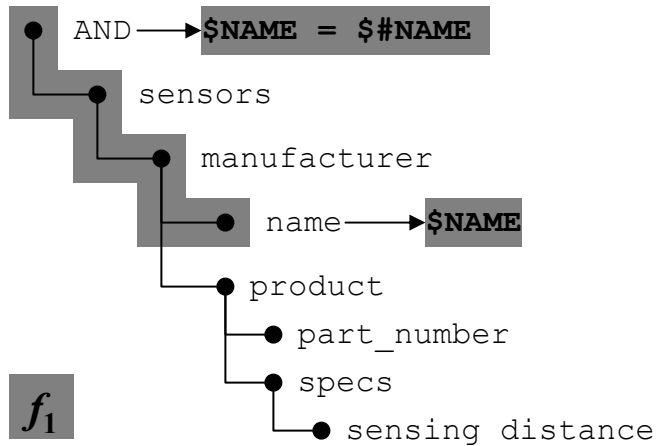
NEMA1

option

QURSED Editor

From visual actions to QSS

`sensors/manufacturer/*`
`= man_name_select`

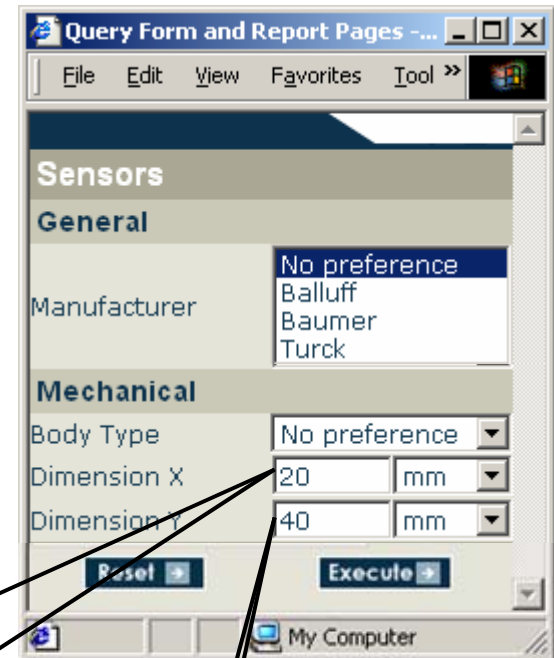
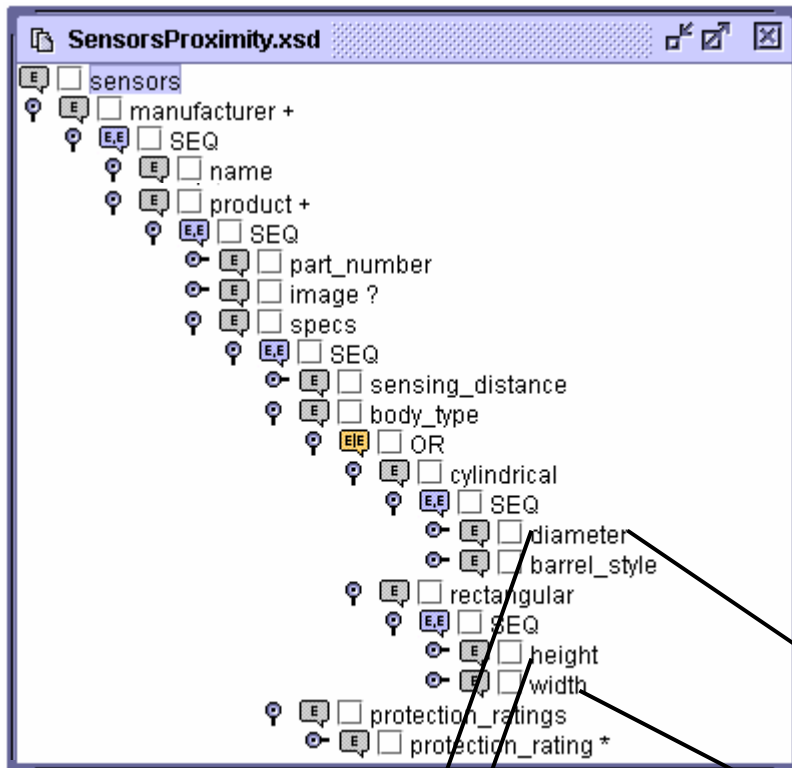


Choice of schema element e means

- Addition of e to the *CTG*
- Addition of the e path to the *CTG*
- Creation of a name variable for e

QURSED Editor

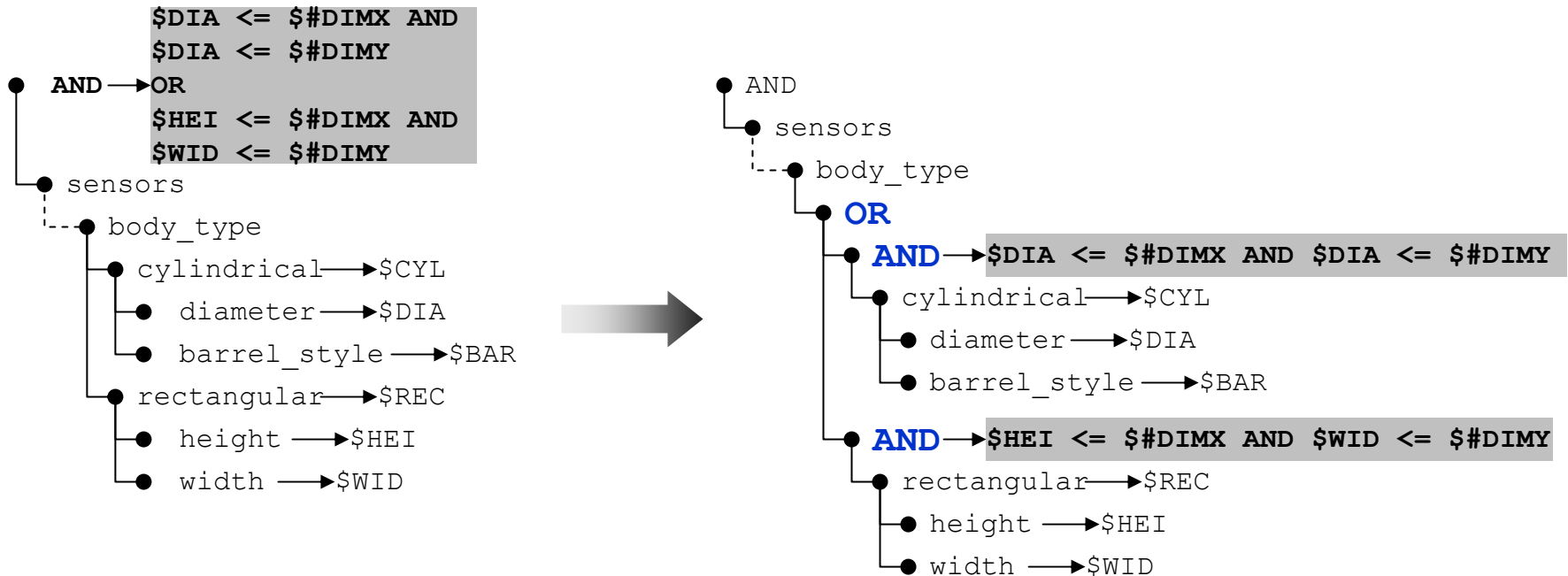
Disjunction



$(\$DIA \leq \$\#DIMX \text{ AND } \$DIA \leq \$\#DIMY)$
OR
 $(\$HEI \leq \$\#DIMX \text{ AND } \$WID \leq \$\#DIMY)$

QURSED Editor

Disjunction



- Creation of disjunctive condition triggers
transformation of the Condition Tree Generator
– **ORNodes Algorithm**

QURSED Editor

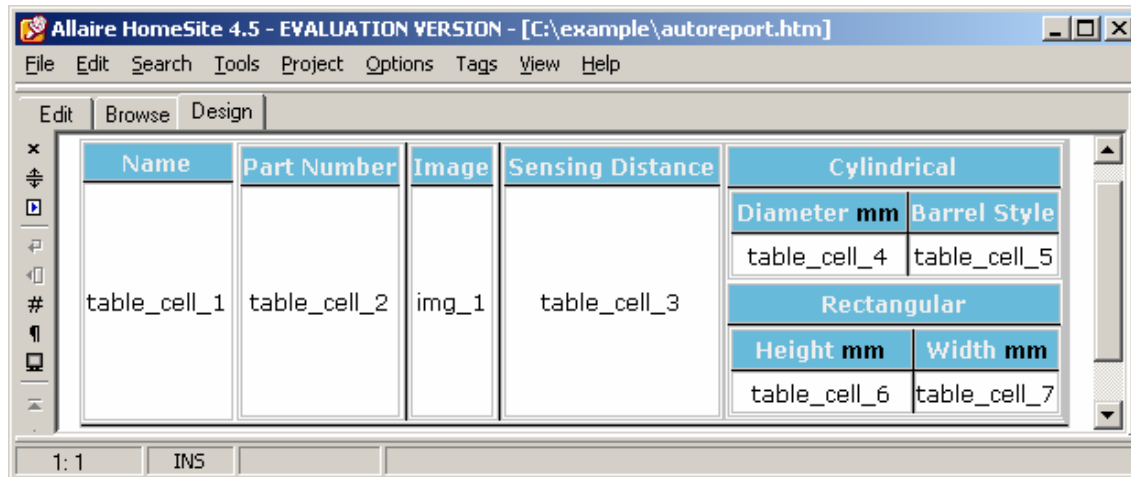
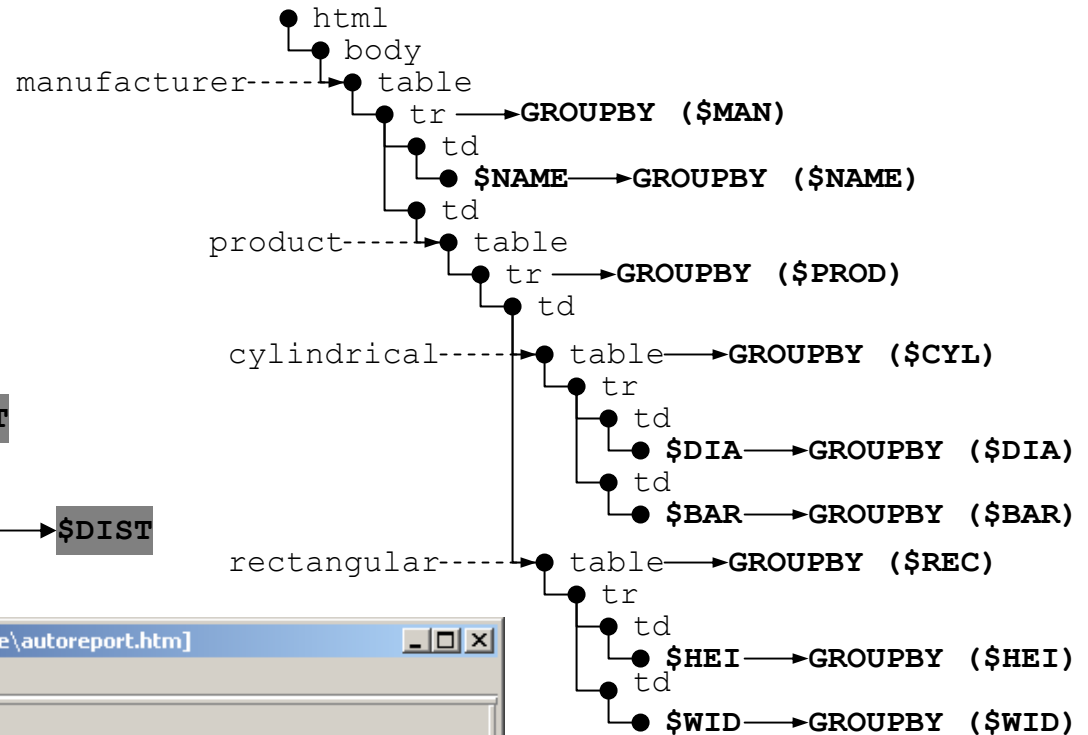
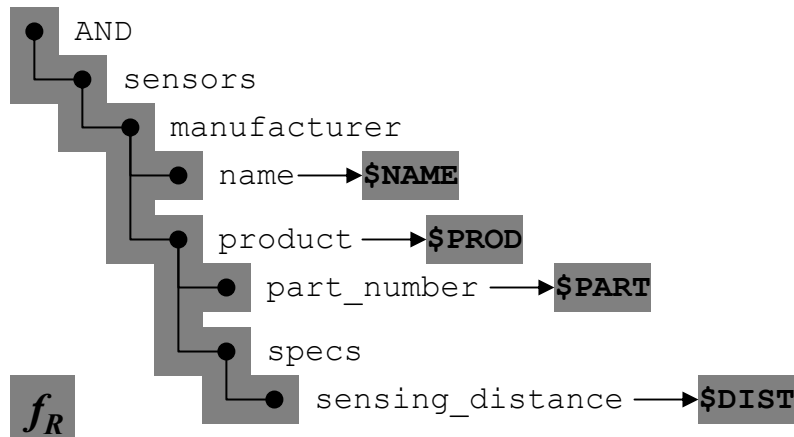
Building Reports

The screenshot displays the QURSED Editor interface, which is used for building reports from data sources. The interface is divided into several panes:

- Data Source(s):** Shows the XML structure of the data source, `SensorsProximity.xsd`. The tree includes elements like `sensors`, `manufacturer`, `product`, `part_number`, `image`, `specs`, `sensing_distance`, `body_type`, `cylindrical`, `diameter`, `barrel_style`, and `rectangular`. A blue box labeled "Elements to Appear on Report" highlights the `manufacturer`, `product`, and `body_type` elements.
- Query Form:** Contains three mapping tables:
 - Element Mappings:** Maps source elements to report targets. For example, `/manufacturer/name*` maps to `table_cell_1`, `.../part_number*` maps to `table_cell_2`, `.../product/image*` maps to `img_1`, `.../sensing_distance*` maps to `table_cell_3`, and `/diameter*` maps to `table_cell_4`.
 - GroupBy Mappings:** Maps source elements to report targets. For example, `sensors/manufacturer` maps to `table_row_1`, `.../manufacturer/prod...` maps to `table_row_2`, `.../product/image*` maps to `img_1`, `.../body_type/cylindrical` maps to `table_row_3`, and `/body_type/rectangu` maps to `table_row_4`.
 - SortBy Mappings:** Currently empty.
- Report:** Shows the resulting report structure, including `html`, `body`, `table`, `tr`, `td`, and `img` elements. A blue box labeled "Group By Mapping" points to the `tr` element, and another blue box labeled "Element Mapping" points to the `td` element.

QURSED Editor

Result Tree



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 QURSED 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

?