

# **QURSED: Querying and Reporting Semistructured Data**

---

**Yannis Papakonstantinou**  
**Michalis Petropoulos**

UNIVERSITY OF CALIFORNIA, SAN DIEGO

**Vasilis Vassalos**

NEW YORK UNIVERSITY

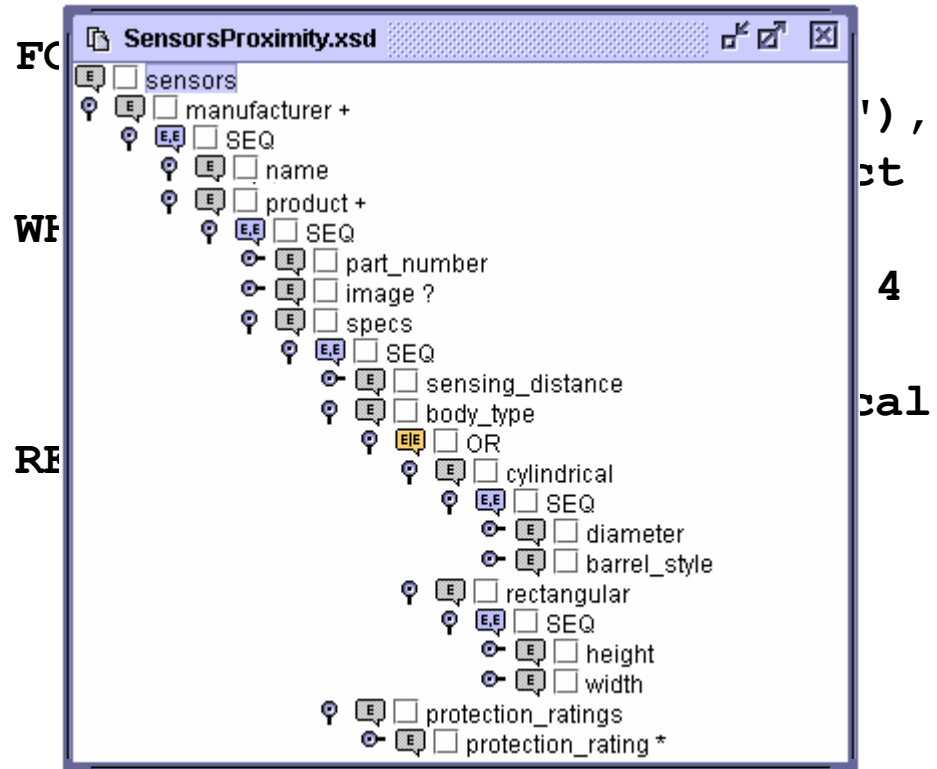
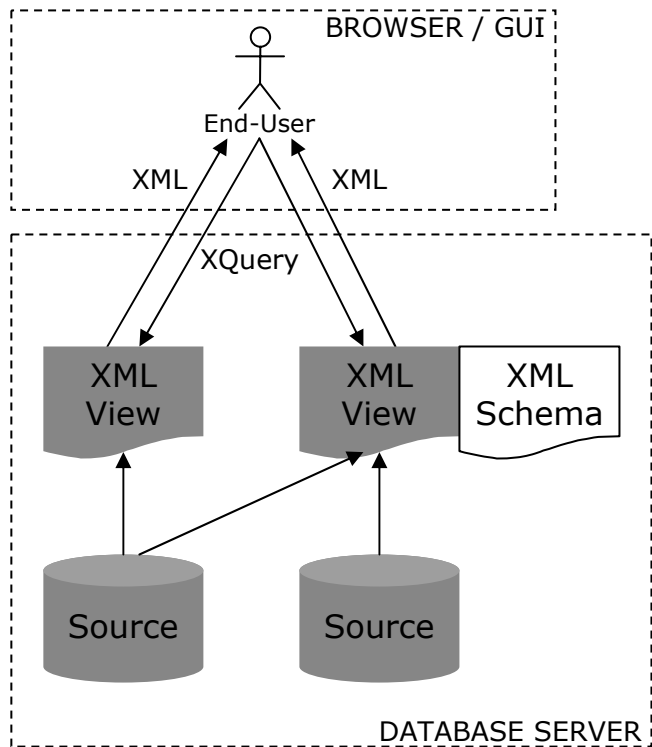
June 2002

# Overview

---

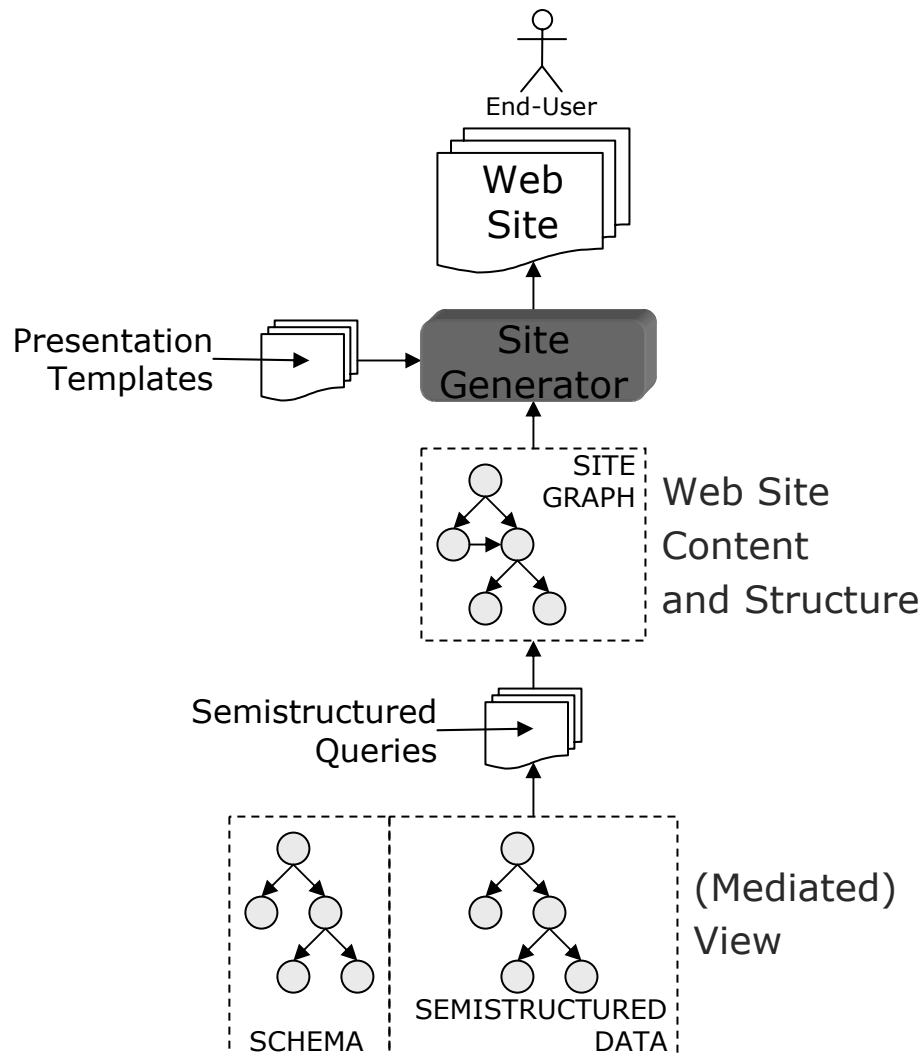
- Query Forms and Reports
  - Challenges of Semistructured Data
- The QURSED system
  - Architecture
- Technical foundation
  - Tree Query Language (TQL)
  - Query Set Specification (QSS)
- QURSED Editor

# 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 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:

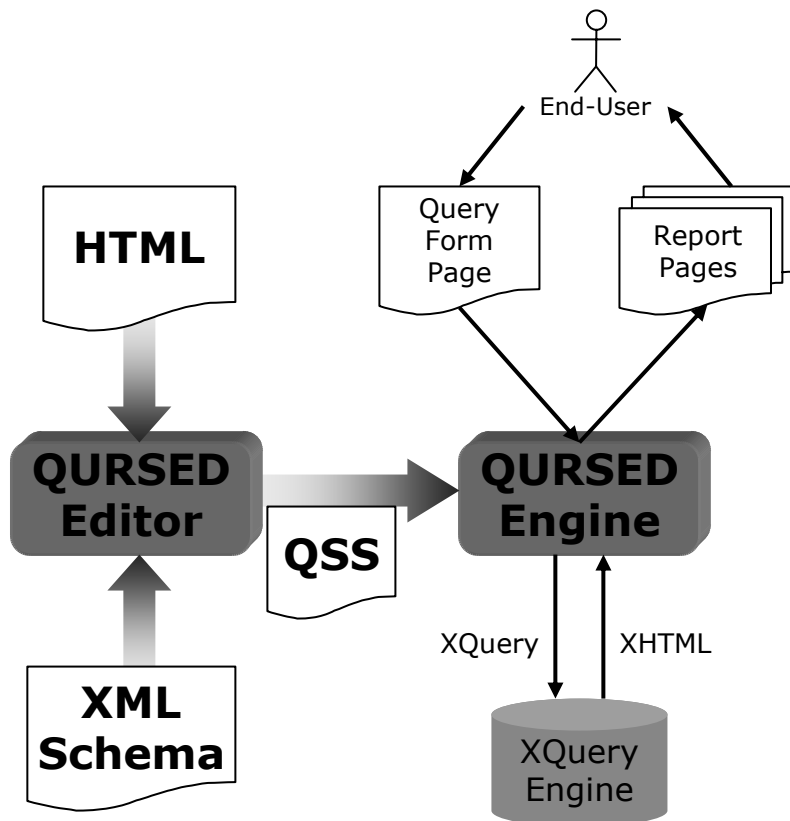
Sort By Options: Sensing Distance 
  
DESC-Manufacturer

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

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

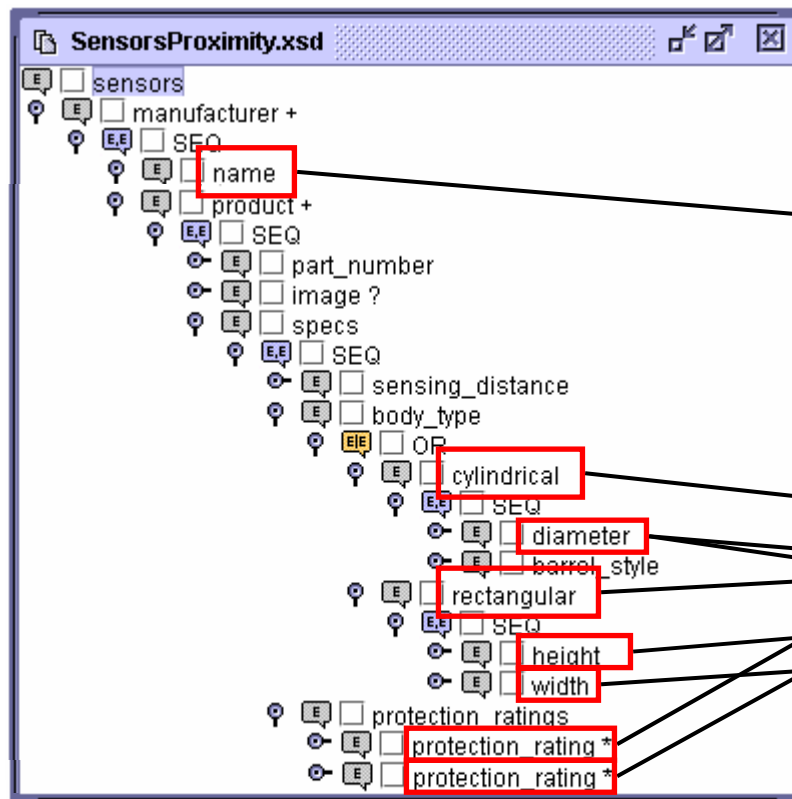
## QURSED Editor

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.
- 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 several `select` elements. The first `select` element is selected, showing `man_name_select` as the selected option. Other options include `Baumer` and `Turck`. The second `select` element is also selected, showing `prot_rating_1_select` as the selected option. Other options include `NEMA3` and `NEMA1`.



# Developing Query Forms from the XML Schema

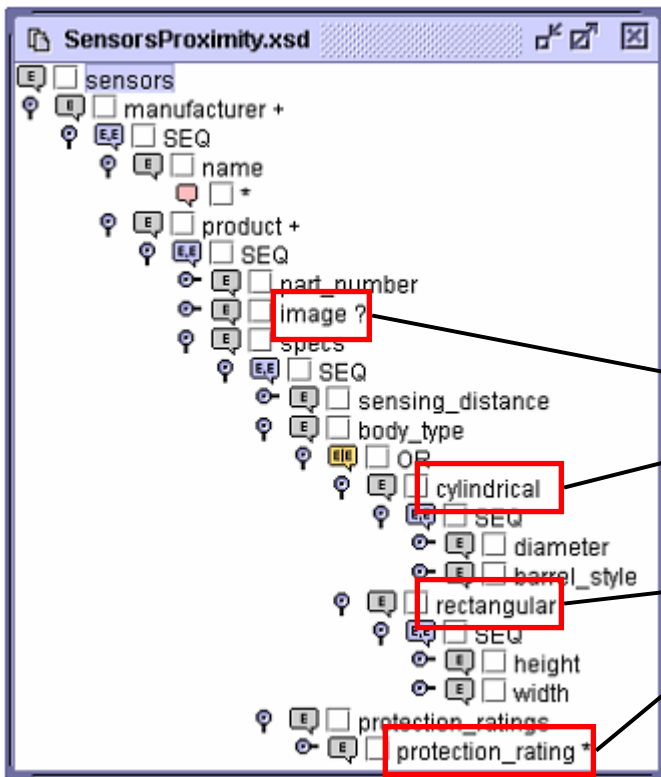


The screenshot shows a web application titled "Query Form and Report Pages". The form is for "Sensors" and is divided into several sections:

- General:** Contains fields for "Manufacturer" (a dropdown menu with "Turck" selected), "Sensing Distance" (a text input with "6" and a "mm" dropdown), "Protection Rating" (a dropdown menu with "No preference" selected), and "Protection Rating" (another dropdown menu with "No preference" selected).
- Mechanical:** Contains fields for "Body Type" (a dropdown menu with "Rectangular" selected), "Height" (a text input with "mm" dropdown), and "Width" (a text input with "mm" dropdown).
- Results:** Contains fields for "Results/page" (a dropdown menu with "10" selected), "Sort By Options" (a dropdown menu with "Sensing Distance" selected and "DESC" dropdown), and "DESC-Manufacturer" (a dropdown menu with "up" and "dn" buttons).

Buttons for "Reset" and "Execute" are located at the bottom of the form. The Windows taskbar at the bottom shows "My Computer".






# Developing Reports from the XML Schema



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

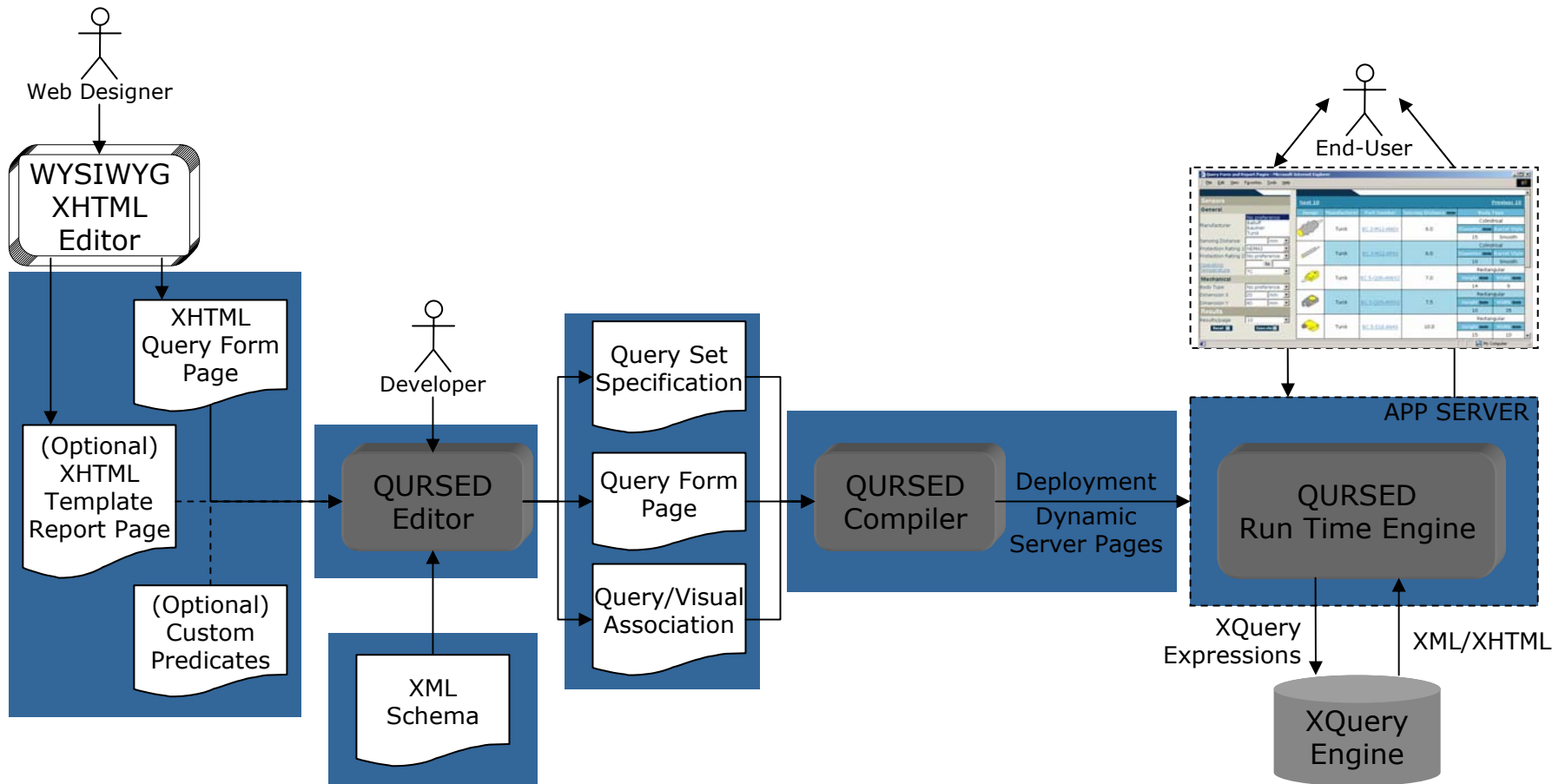
File Edit View Favorites Tools Help

Next 10 Previous 10

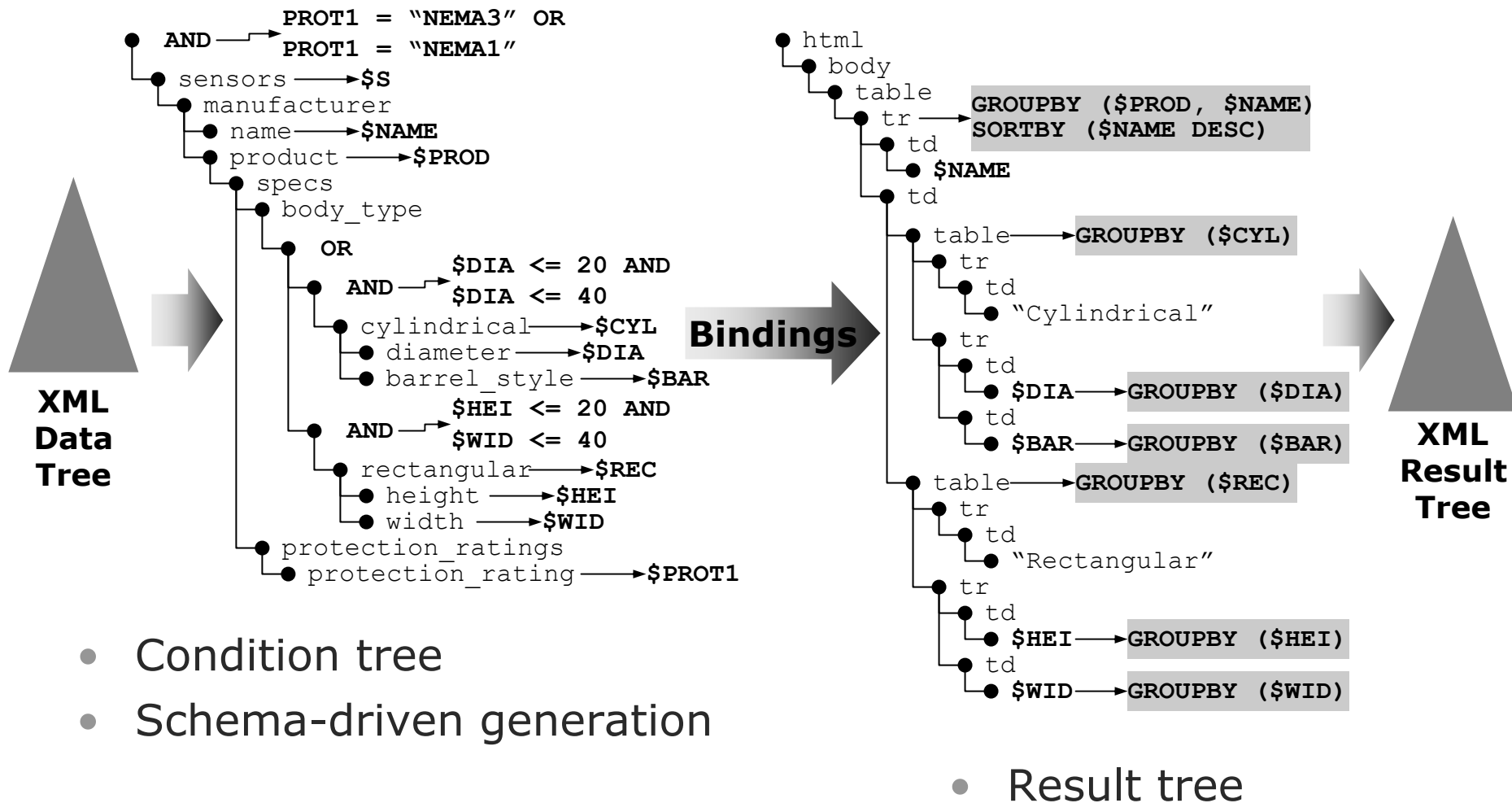
Image	Manufacturer	Part Number	Protection Ratings	Sensing Distance mm	Body Type
	Turck	<a href="#">BC 3-M12-AN6X</a>	NEMA1 NEMA3 NEMA4	6.0	Cylindrical Diameter mm Barrel Style 15 Smooth
	Turck	<a href="#">BC 3-M12-AP6X</a>	NEMA3	6.0	Cylindrical Diameter mm Barrel Style 19 Smooth
	Turck	<a href="#">BC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

My Computer

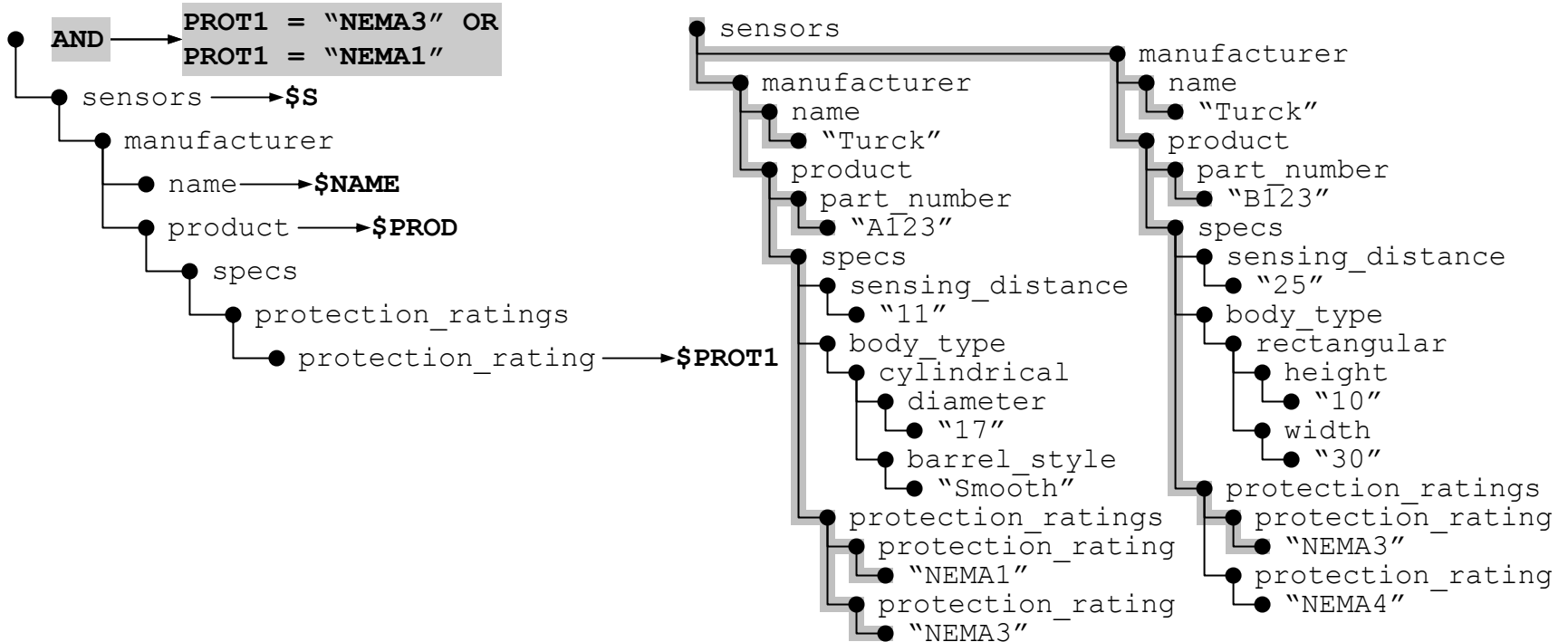
# QRUSED System Architecture



# Tree Query Language (TQL)



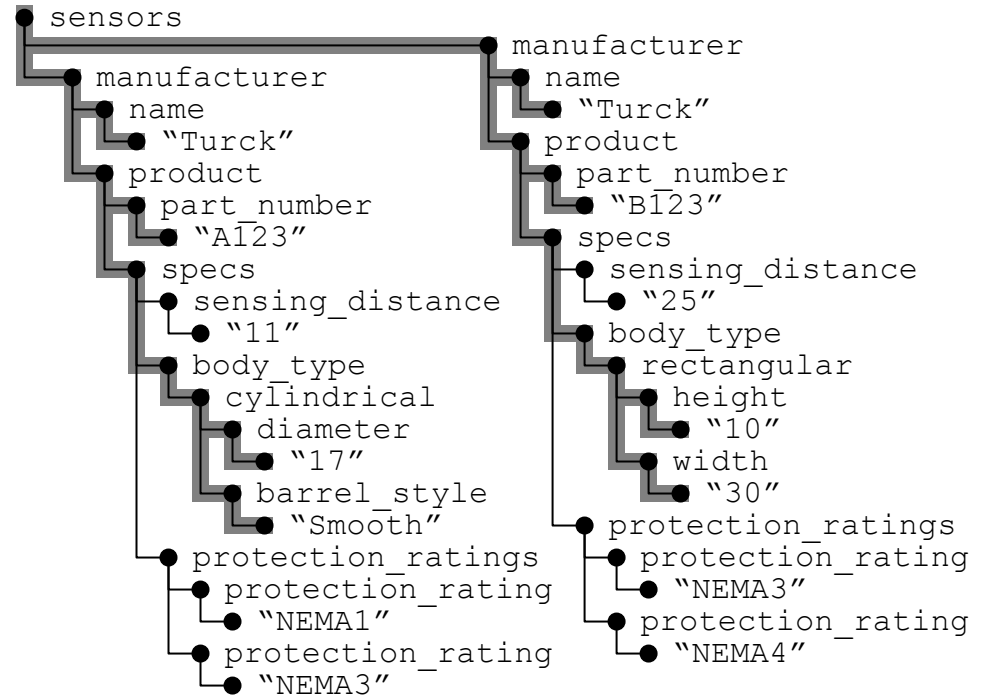
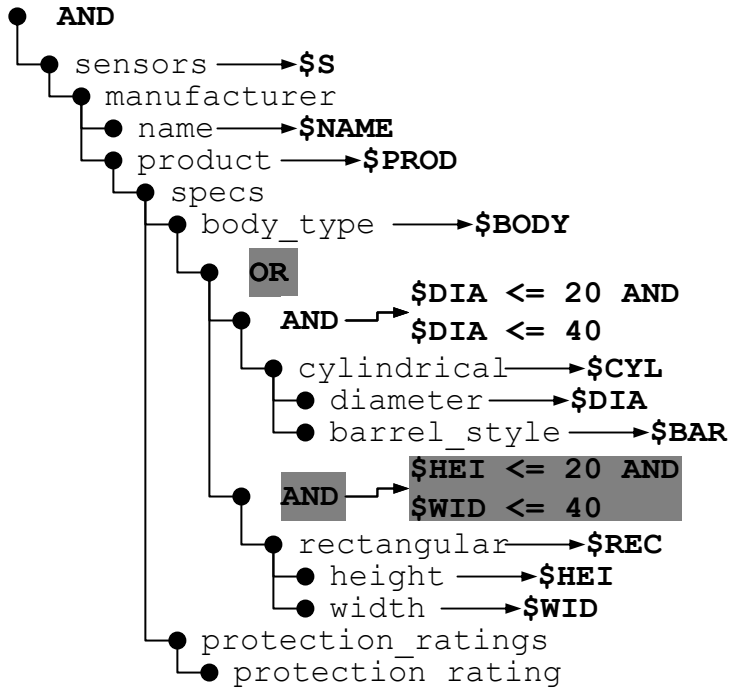
# Tree Query Language (TQL)



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



# Tree Query Language (TQL)

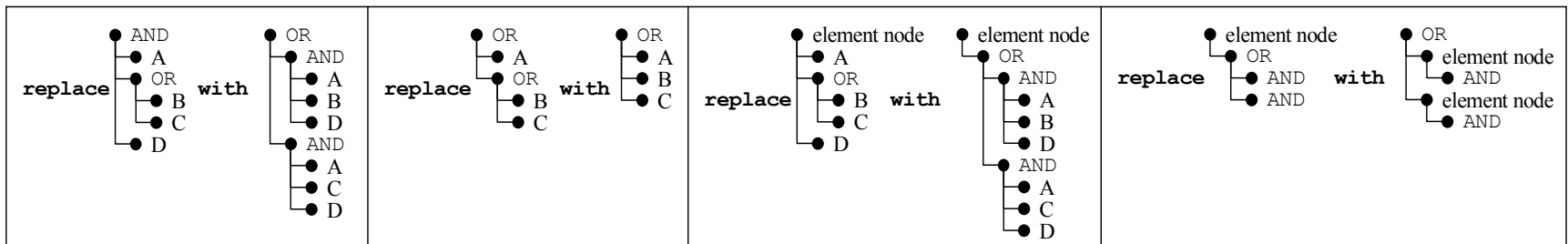


\$NAME	\$PROD	\$PART	\$BODY	\$CYL	\$DIA	\$BAR	\$REC	\$HEI	\$WID
Turck	<ul style="list-style-type: none"> <li>product</li> <li>part_number</li> <li>• "A123"</li> </ul>	A123	cylindrical	<ul style="list-style-type: none"> <li>cylindrical</li> <li>diameter</li> <li>• "17"</li> </ul>	17	Smooth			
Turck	<ul style="list-style-type: none"> <li>product</li> <li>part_number</li> <li>• "B123"</li> </ul>	B123	rectangular				<ul style="list-style-type: none"> <li>rectangular</li> <li>height</li> <li>• "10"</li> </ul>	10	30

# TQL Semantics

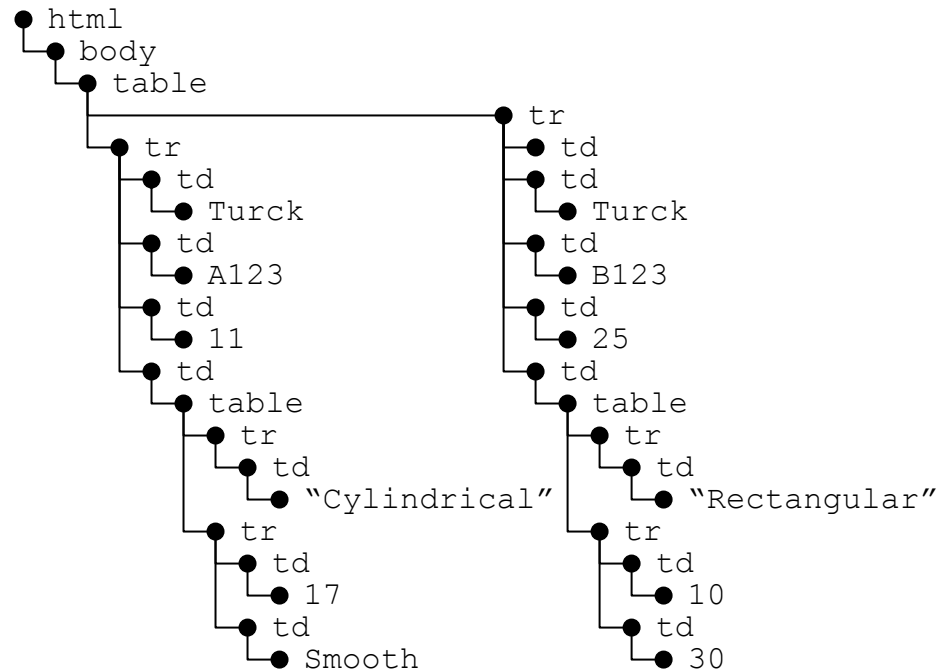
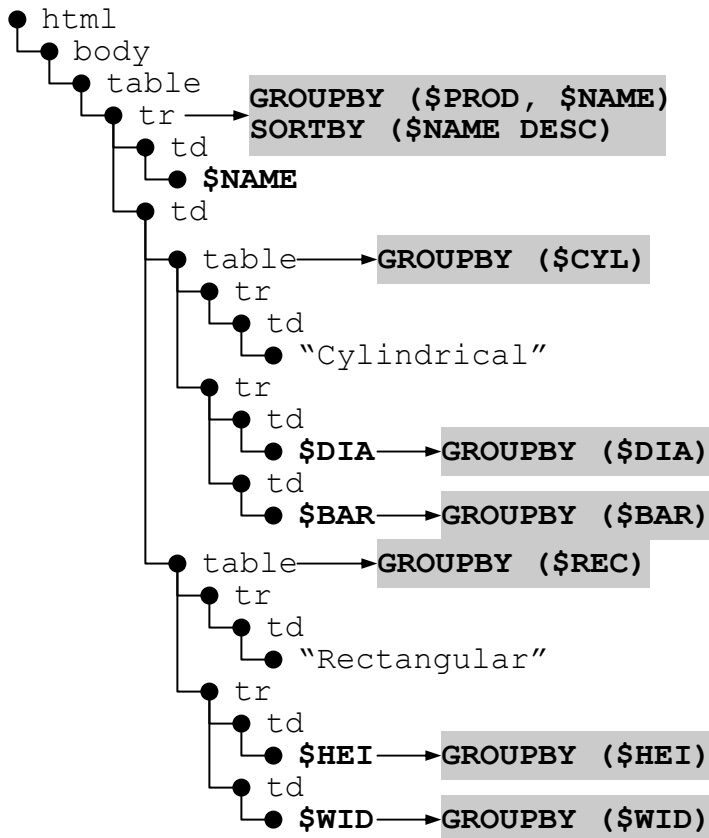
## Condition Tree


- 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

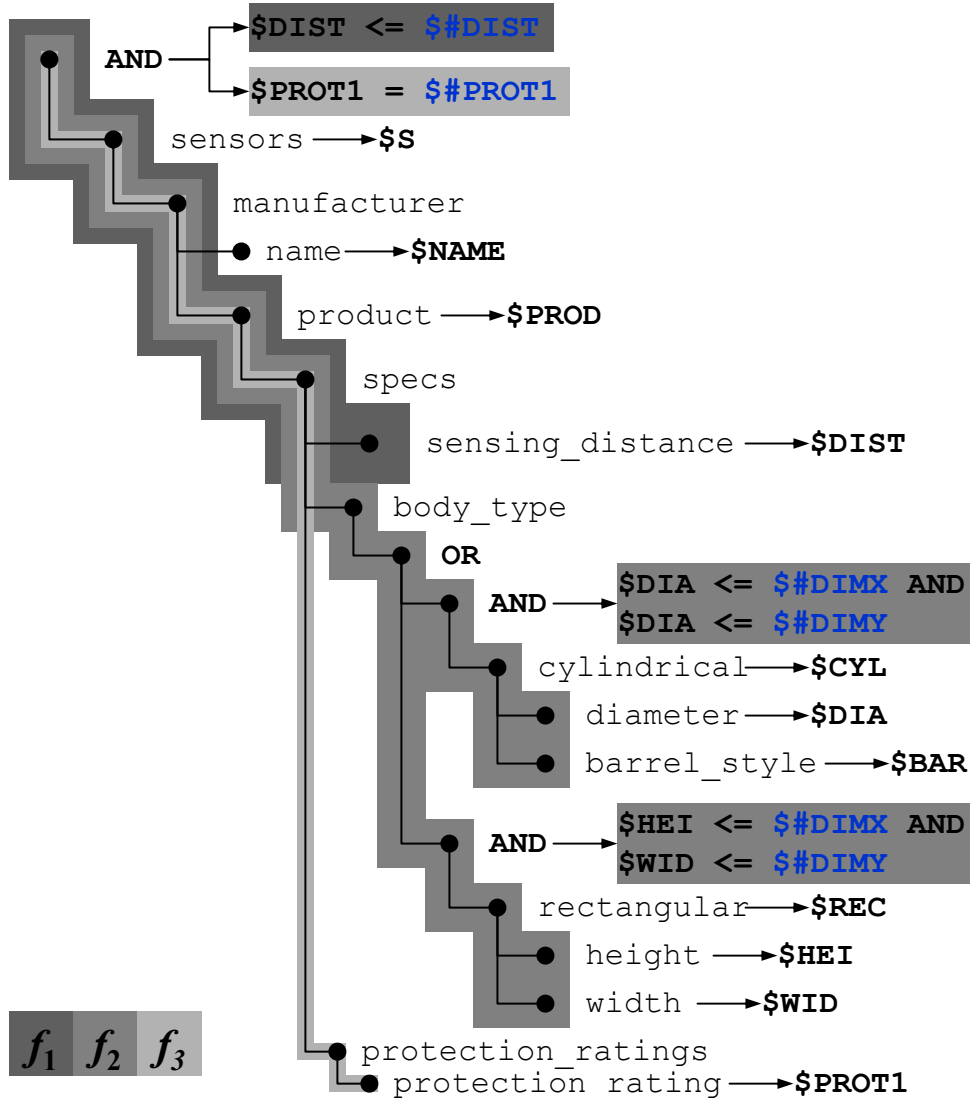


# Tree Query Language (TQL)

---

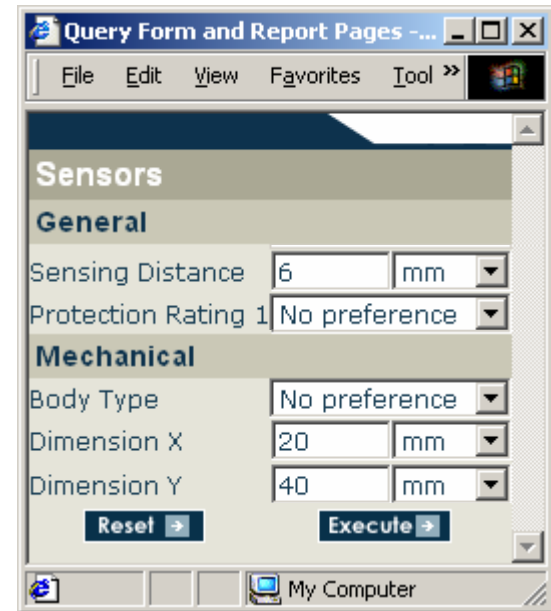
- Translated to XQuery
  - By QURSED Run-Time Engine
  - **TQL2XQuery Algorithm**
  - Syntax directed translation
  - Tree patterns in TQL to nested FOR-WHERE-RETURN expressions in XQuery

# Query Set Specification (QSS)

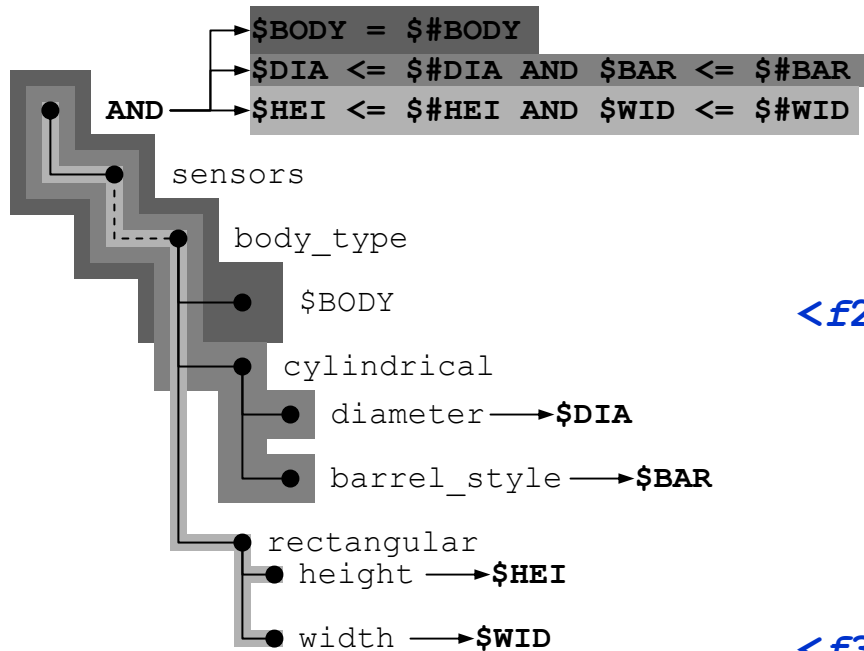


## Condition Tree Generator

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



# Dependencies



Mechanical	
Body Type	Cylindrical
Diameter	mm
Barrel Style	No preference

$\langle f_2, \text{\$}\#BODY = \text{"Cylindrical"}, \{f_1\} \rangle$

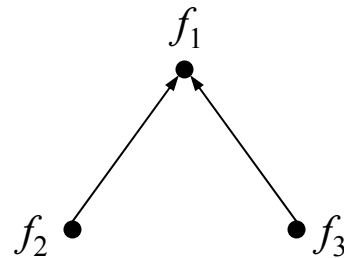
Mechanical	
Body Type	Rectangular
Height	mm
Width	mm

$\langle f_3, \text{\$}\#BODY = \text{"Rectangular"}, \{f_1\} \rangle$

$f_1$   $f_2$   $f_3$

# Dependencies

---

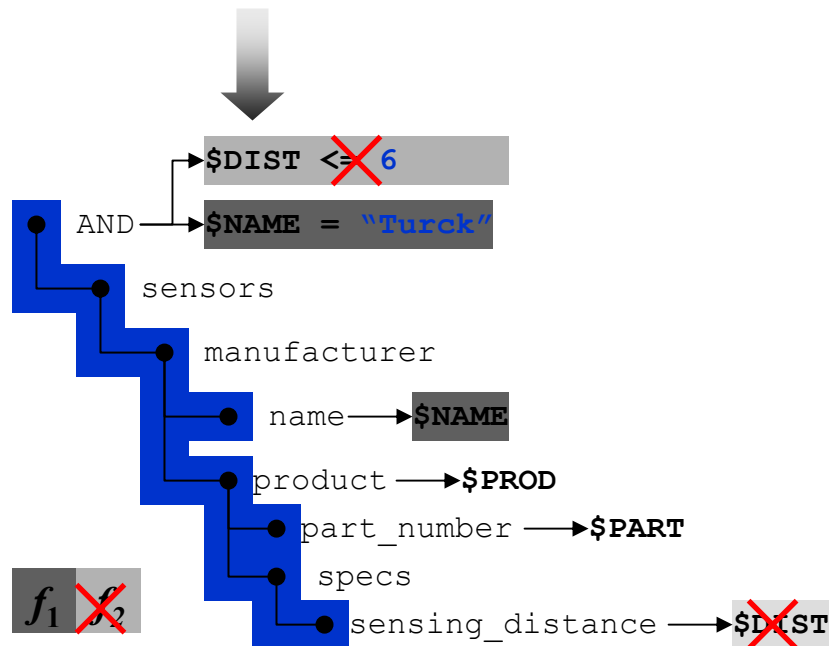
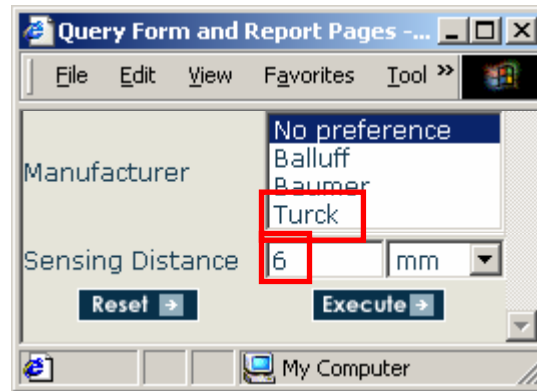


`<f2, $#BODY = "cylindrical", {f1}>`

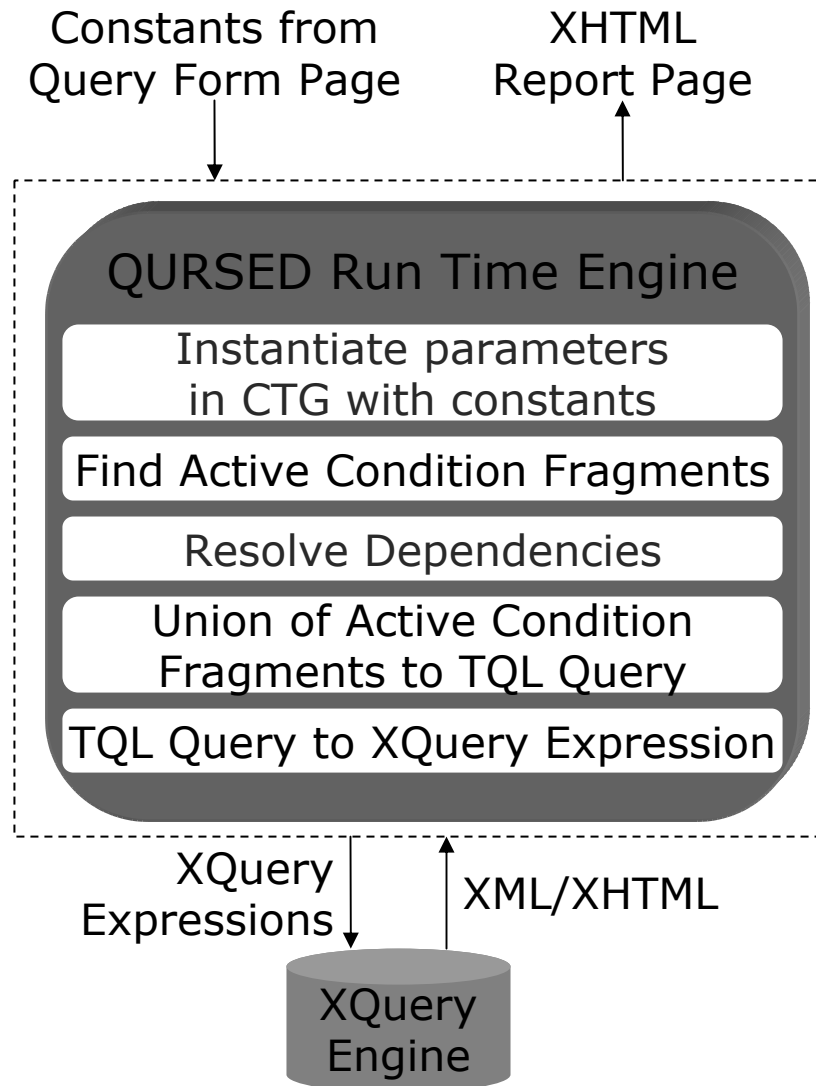
`<f3, $#BODY = "rectangular", {f1}>`

- Dependencies Graph
- Resolution algorithm based on topological sort

# Run-time: QSS to TQL Queries



# Run-time: QSS to TQL Queries



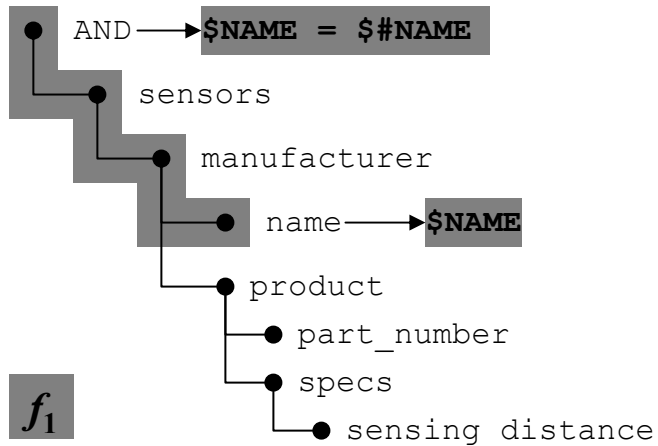
# QURSED Editor

## Building Query/Visual Association

# QURSED Editor

## From visual actions to QSS

`sensors/manufacturer/*`  
`= man_name_select`



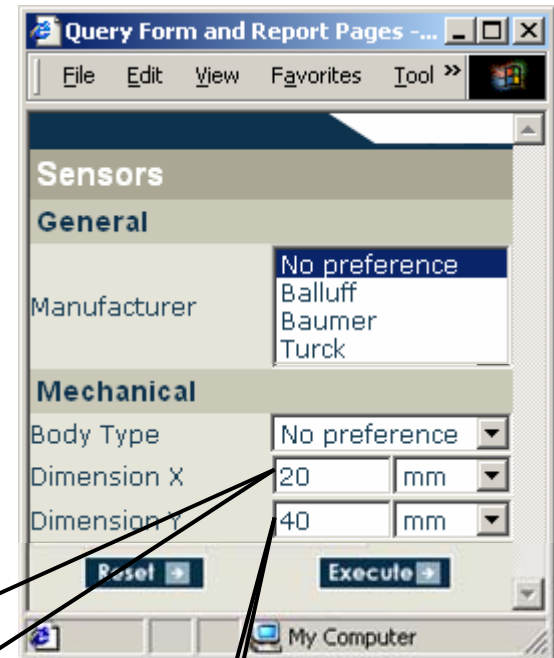
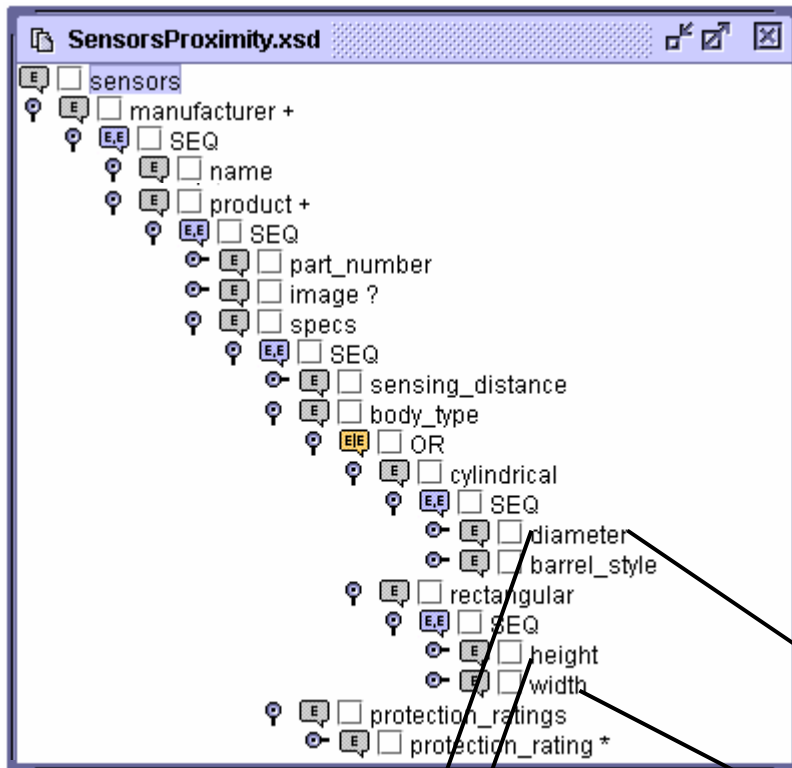
Choice of schema element  $e$  means

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



# QURSED Editor

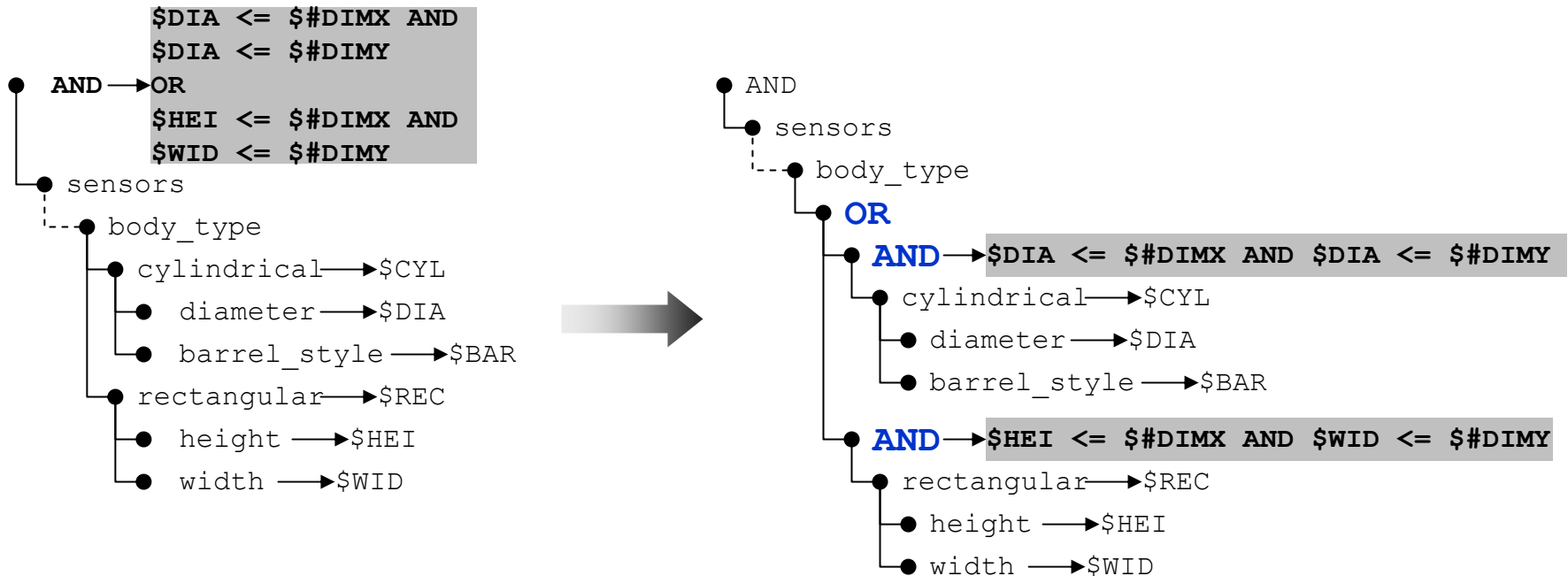
## Disjunction



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

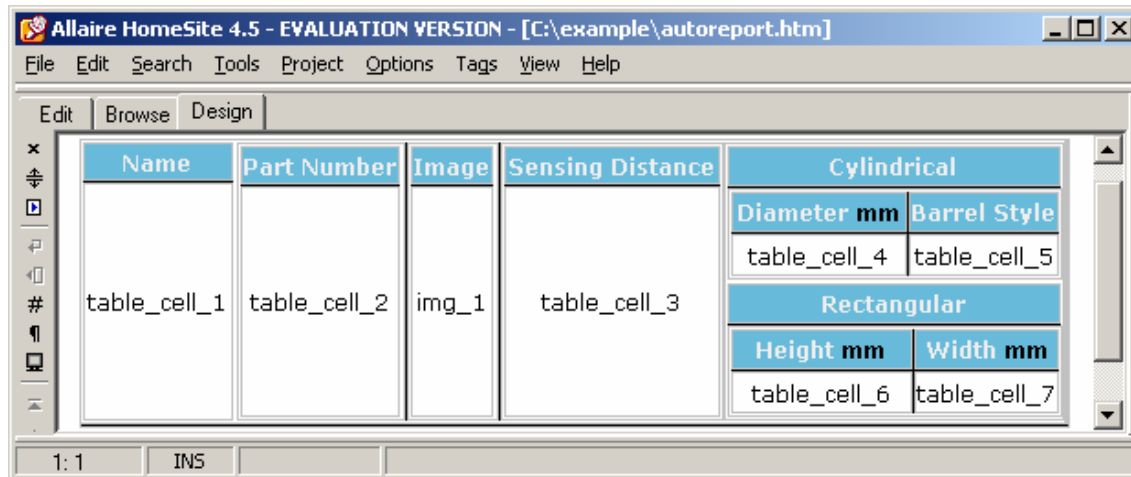
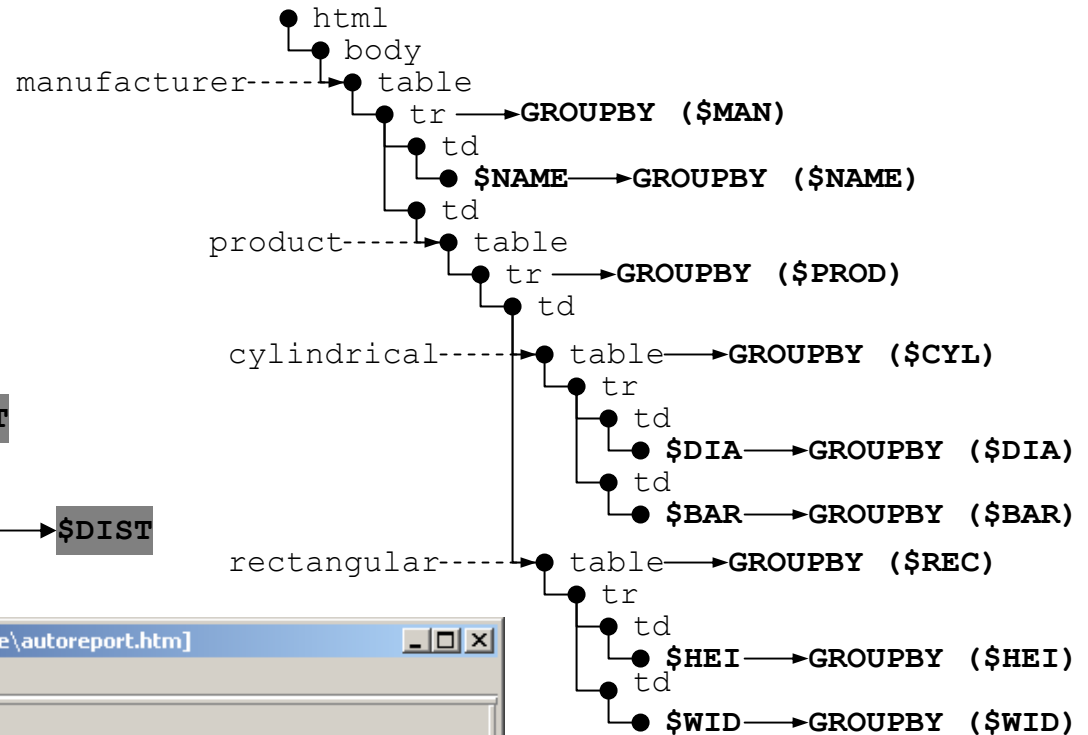
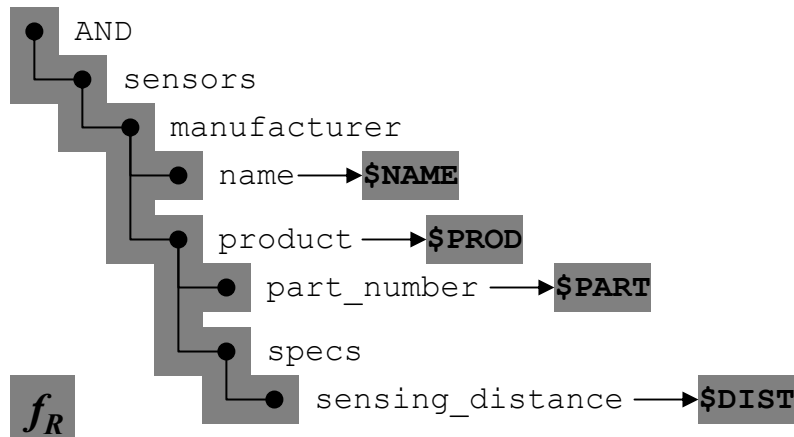
## Building Reports

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

- Data Source(s):** Shows the XML structure of the data source, `SensorsProximity.xsd`. The tree view 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 target report elements. 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 target report elements for grouping. 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 HTML structure of the report. The tree view includes elements like `html`, `body`, `table`, `tr`, `td`, and `img`. Blue arrows point from the "Group By Mapping" and "Element Mapping" tables to the corresponding `tr` and `td` elements in the report tree.

# 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

---

?