

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

---

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

**UNIVERSITY OF CALIFORNIA, SAN DIEGO**

**Vasilis Vassalos**  
**NEW YORK UNIVERSITY**

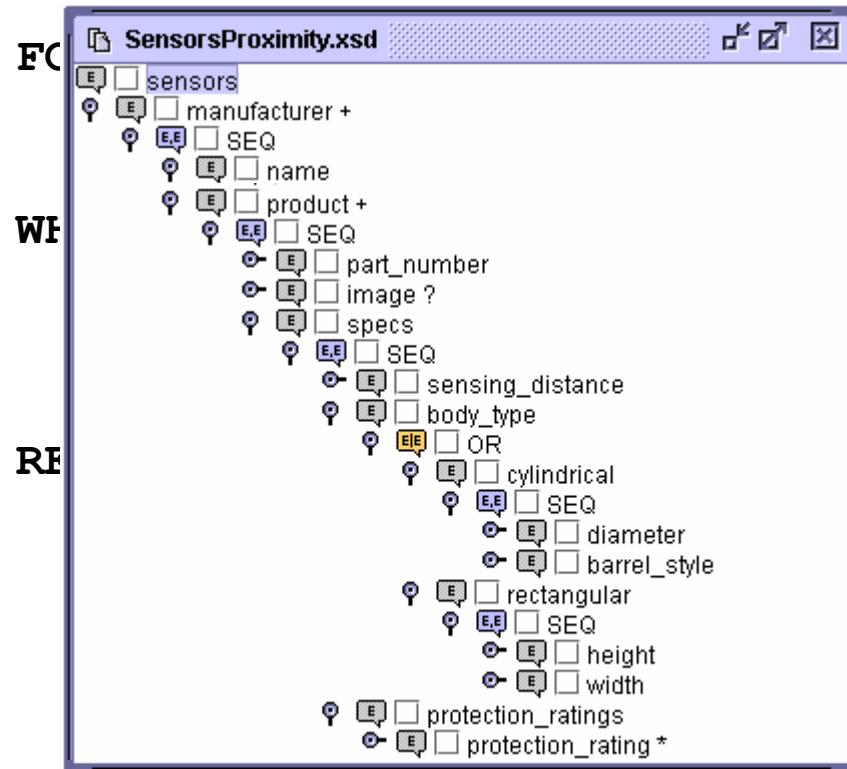
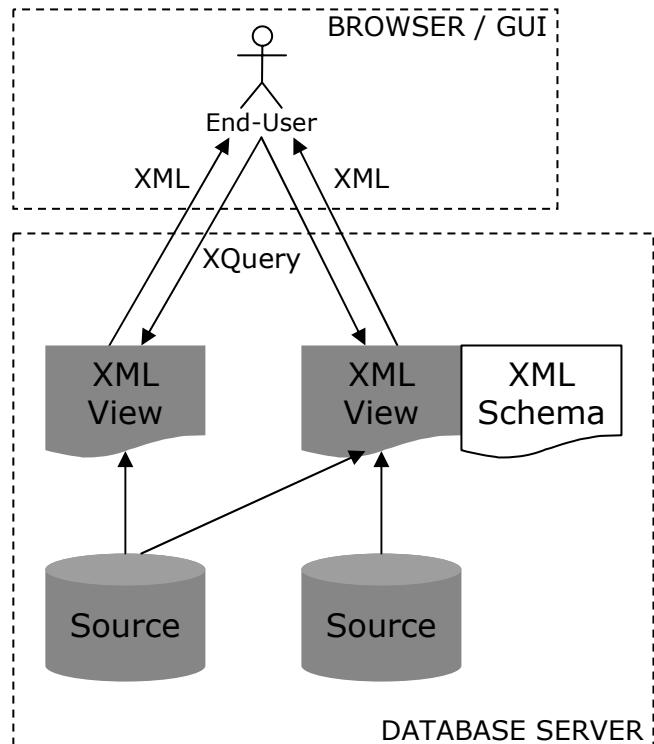
**June 2002**

# Overview

---

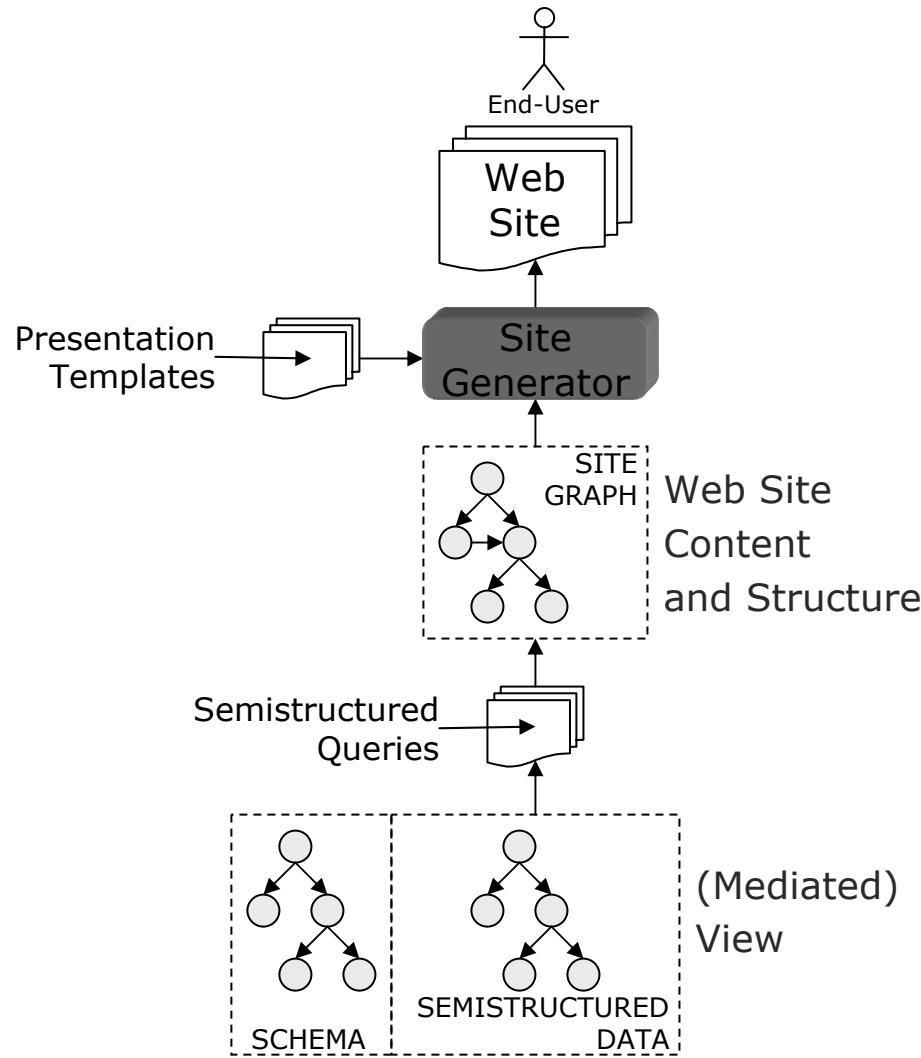
- Query Forms and Reports
  - Challenges of Semistructured Data
- The QURSED system
  - Architecture
- Technical foundation
  - Tree Query Language (TQL)
  - Query Set Specification (QSS)
- QURSED Editor

# 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: No preference, Balluff, Baumer, Turck

Sensing Distance: 6 mm

Protection Rating 1: No preference

Protection Rating 2: No preference

Operating Temperature: to °C

Mechanical

Body Type: No preference

Dimension X: 20 mm

Dimension Y: 40 mm

Results

Results/page: 10

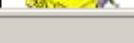
Sort By Options

Sensing Distance: DESC

DESC-Manufacturer: up, dn

Reset  Execute 

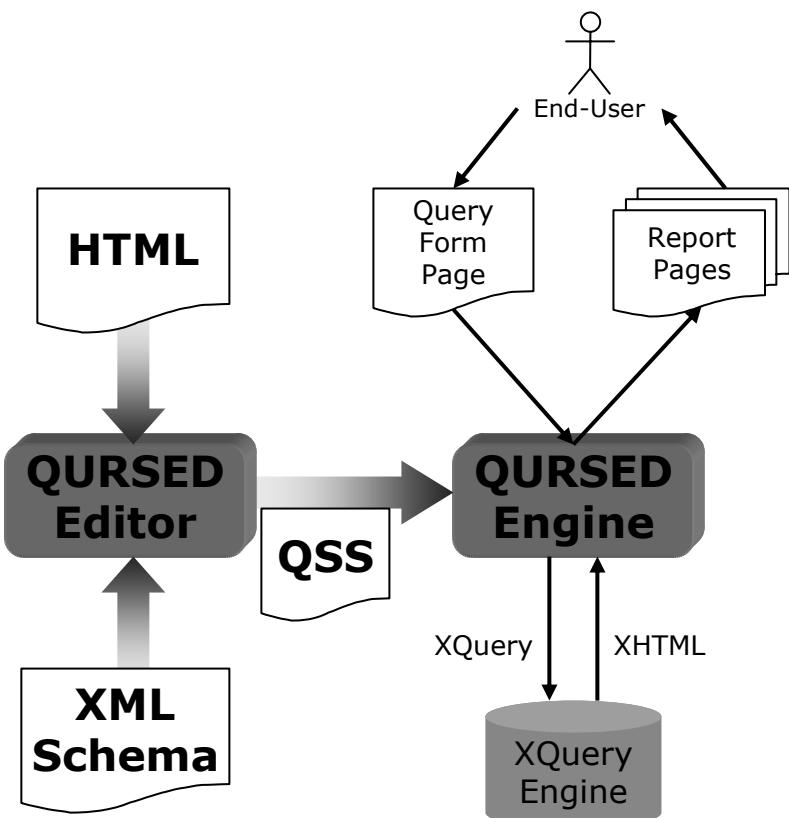
Next 10  Previous 10 

Image	Manufacturer	Part Number	Protection Ratings	Sensing Distance mm	Body Type	
	All		All	All		
	All	BC 3-M12-AN6X	NEMA1	6.0	Cylindrical	
	Balluff		NEMA3		Diameter mm	Barrel Style
	Turck		NEMA4		15	Smooth
	Turck	BC 3-M12-AP6X	NEMA3	6.0	Cylindrical	
					Diameter mm	Barrel Style
					19	Smooth
	Turck	BC 5-Q08-AN6X2	NEMA3	7.0	Rectangular	
					Height mm	Width mm
					14	9
	Turck	BC 5-Q08-AP6X2	NEMA3	7.5	Rectangular	
					Height mm	Width mm
					10	35
	Turck	BC 5-S18-AN4X	NEMA1	10.0	Rectangular	
					Height mm	Width mm
					15	10
	Turck	BC 5-S18-AP4X	NEMA1	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

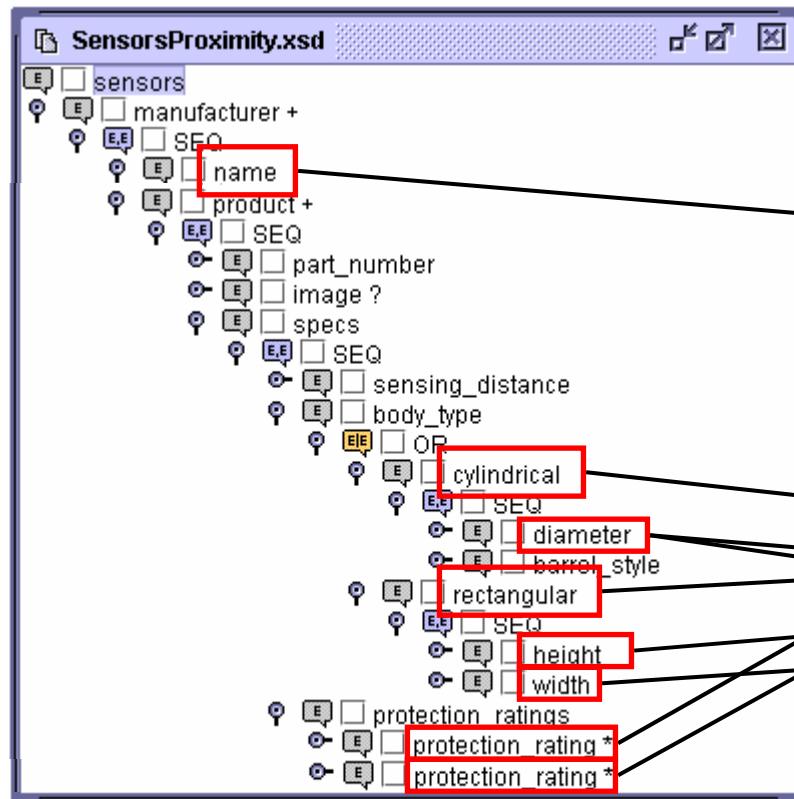
# Query Forms and Reports

## QURSED Editor

The QURSED Editor interface is shown with three main components:

- XML Schema** (Left): Displays the structure of the `SensorsProximity.xsd` XML schema. It shows elements like `sensors`, `manufacturer`, `product`, and various sub-elements and attributes.
- Query Form** (Center): A dialog for creating a Query Form. It includes a **Condition Fragments** section with fields for `manufacturer_name`, `protection_rating_1`, and `protection_rating_2`. Below it is a **Dependencies** section and an **Expression Editor** with tabs for Arithmetic, Comparison, Boolean, Constant, and Custom. The Expression Editor shows a comparison expression: `sensors/manufacturer/name/* = man_name_select`.
- Query Forms and Reports** (Right): A dialog for creating a Query Form Page. It shows the HTML structure: `html body form select name man_name_select option Baumer option Turck select name prot_rating_1_select option NEMA3 option select name prot_rating_2_select option NEMA1 option`.

# Developing Query Forms from the XML Schema



Query Form and Report Pages ...

Sensors

General

Manufacturer: Turck

Sensing Distance: 6 mm

Protection Rating: No preference

Protection Rating: No preference

Operating Temperature: to °C

Mechanical

Body Type: Rectangular

Height: mm

Width: mm

Results

Results/page: 10

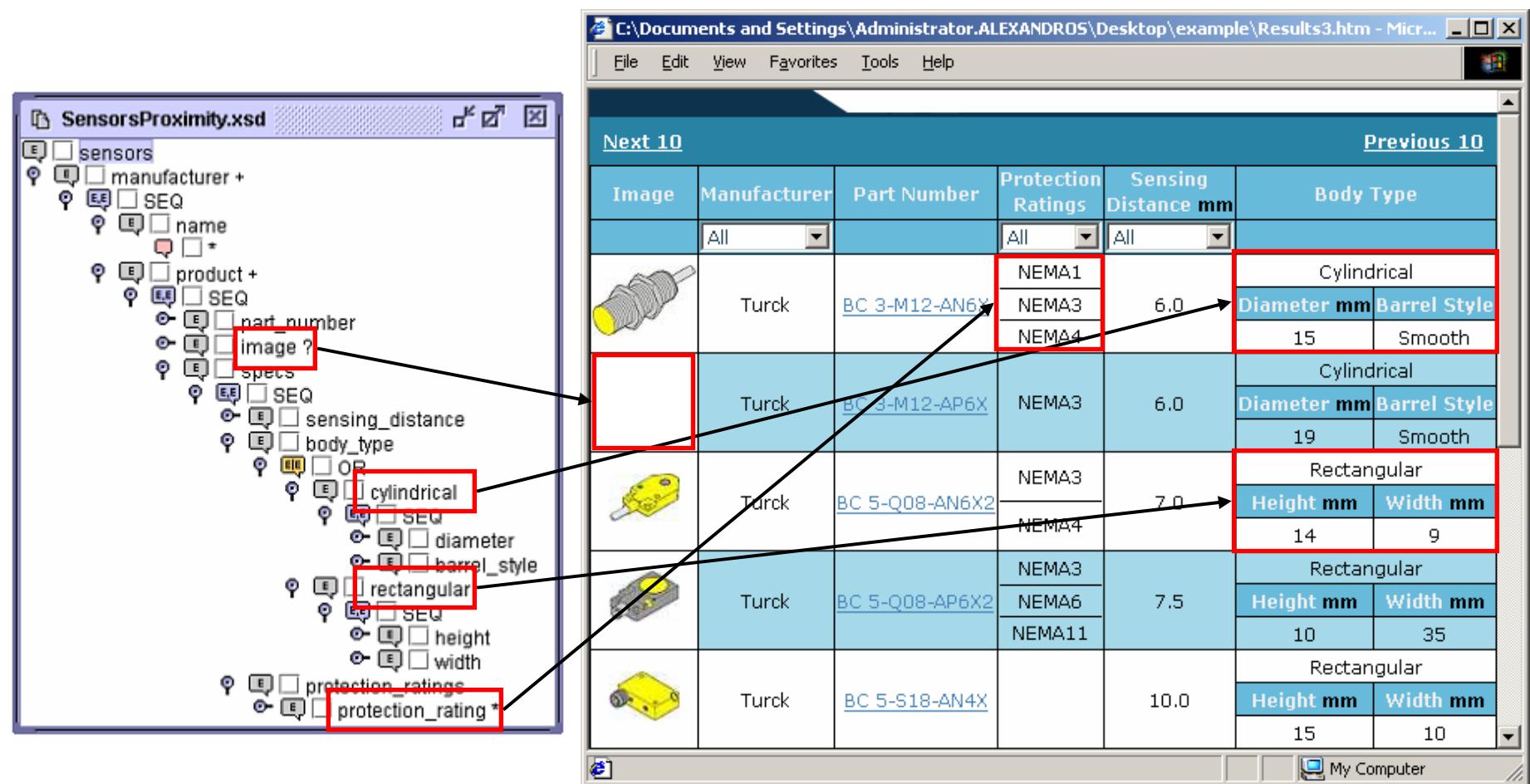
Sort By Options: Sensing Distance DESC

DESC-Manufacturer: up dn

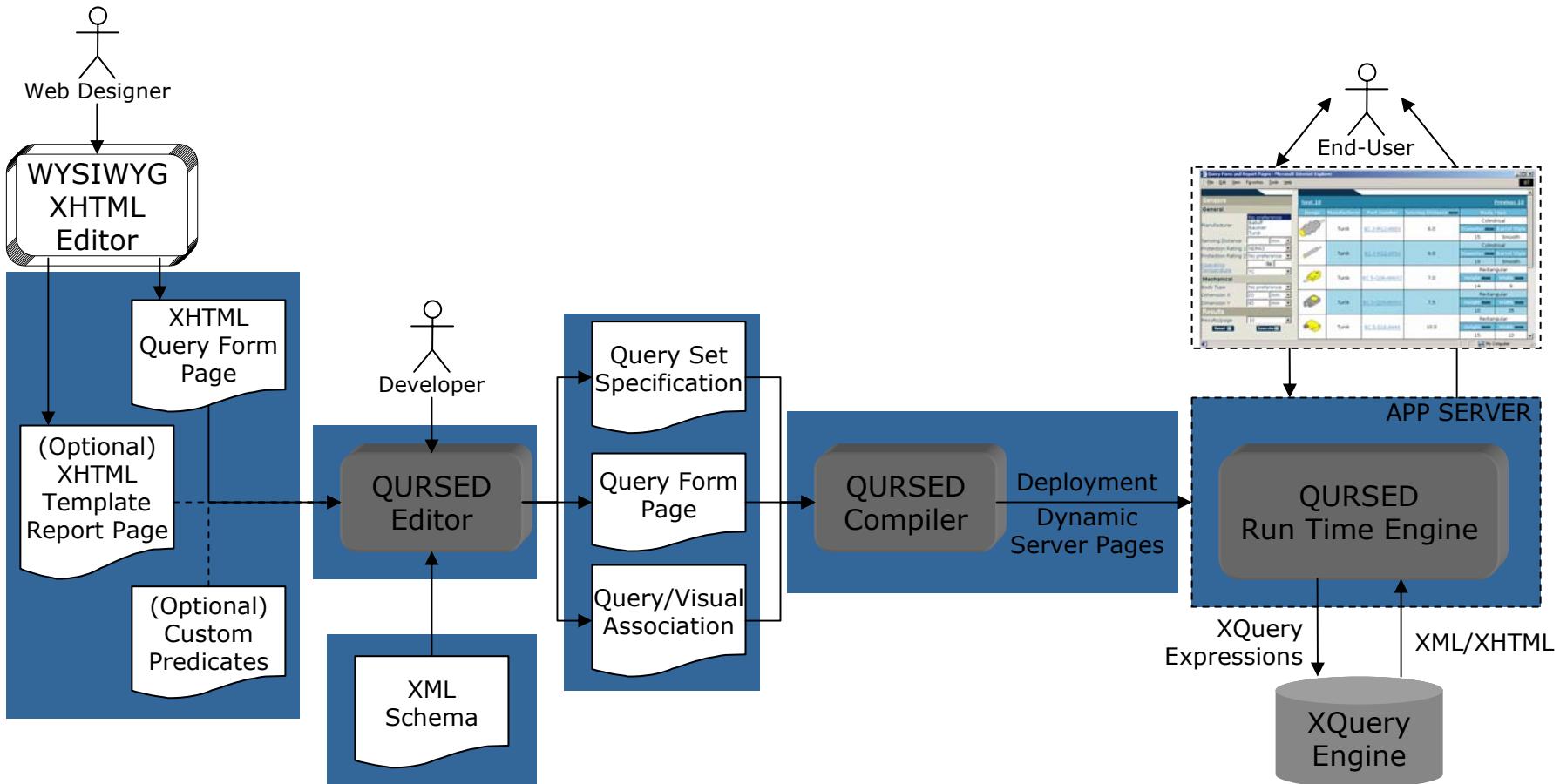
Reset → Execute →

My Computer

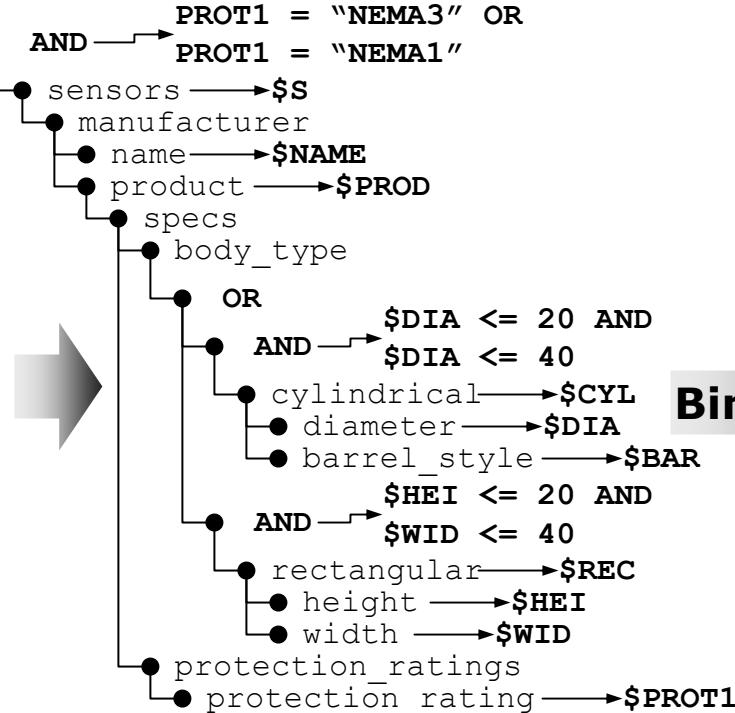
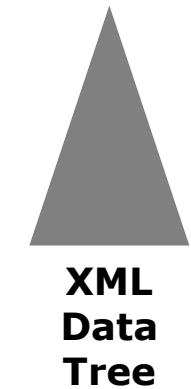
# Developing Reports from the XML Schema



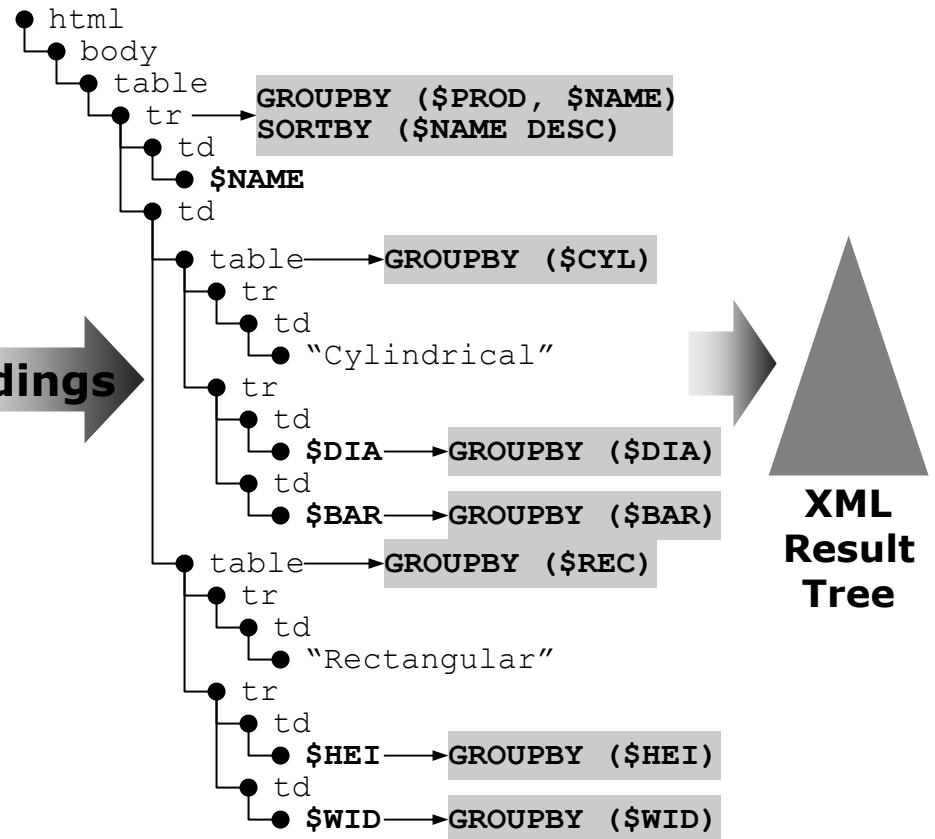
# QURSED System Architecture



# Tree Query Language (TQL)



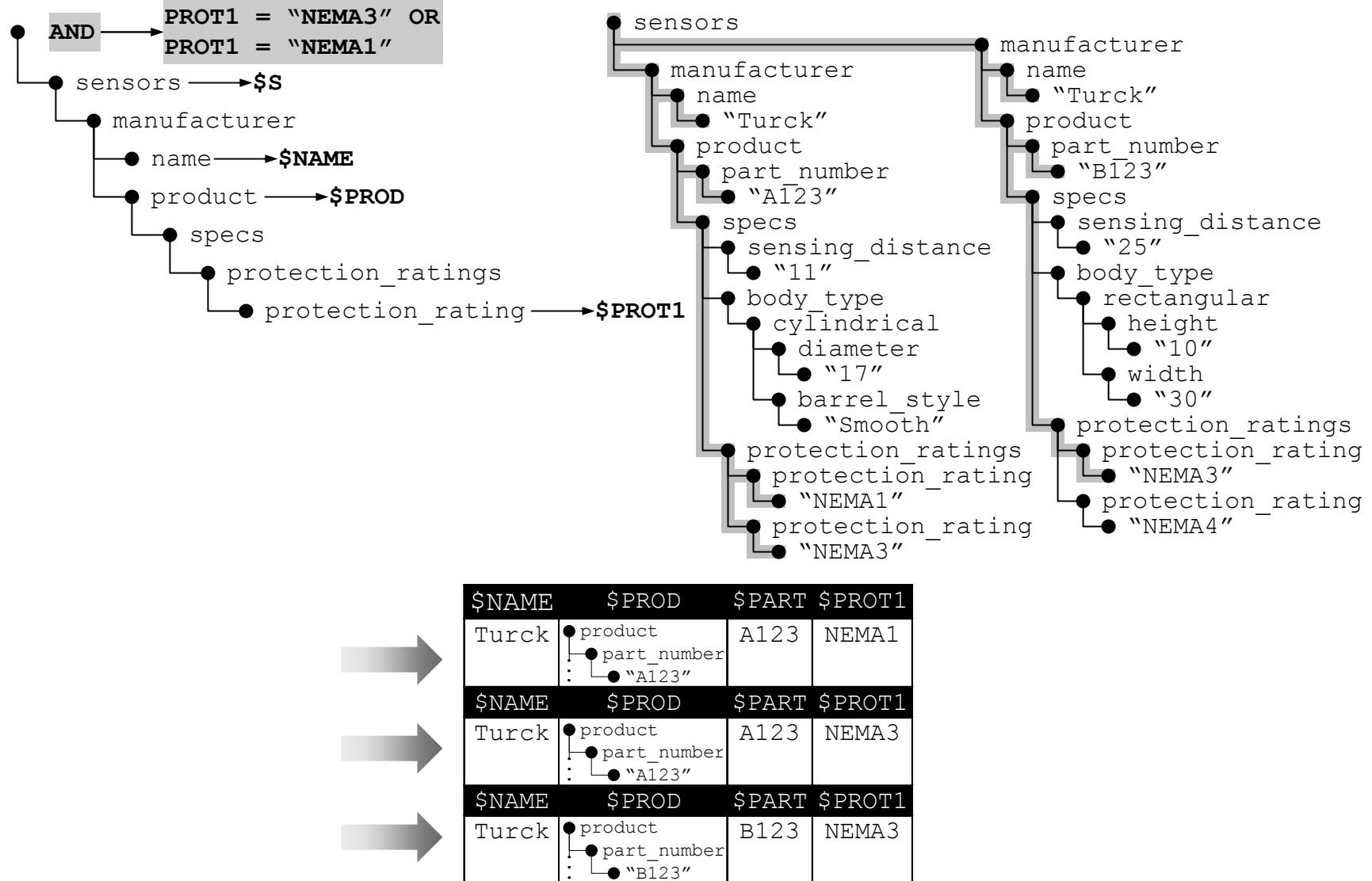
Bindings



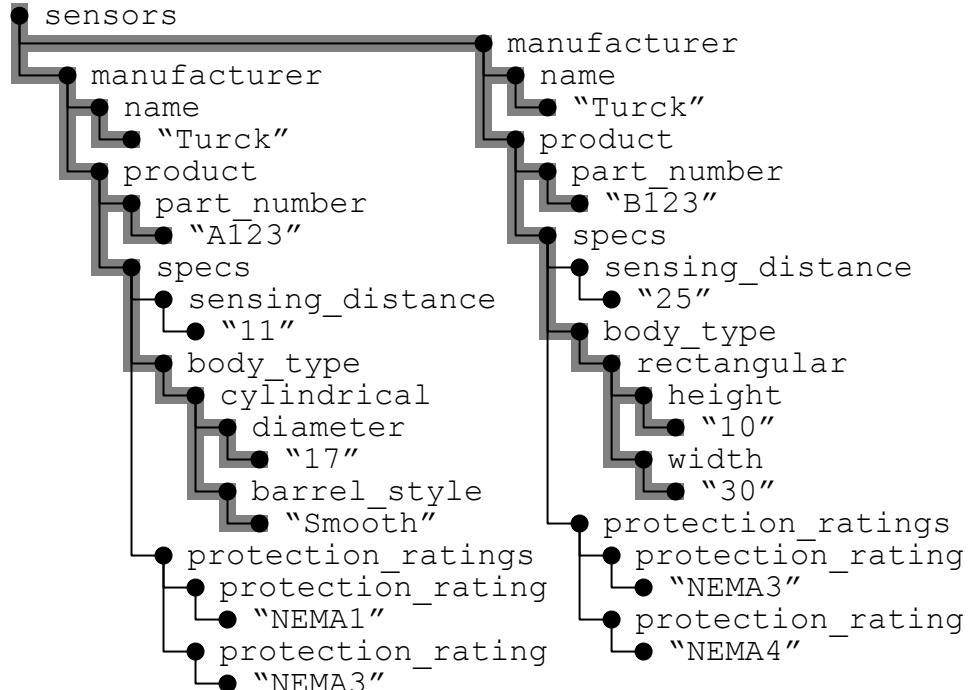
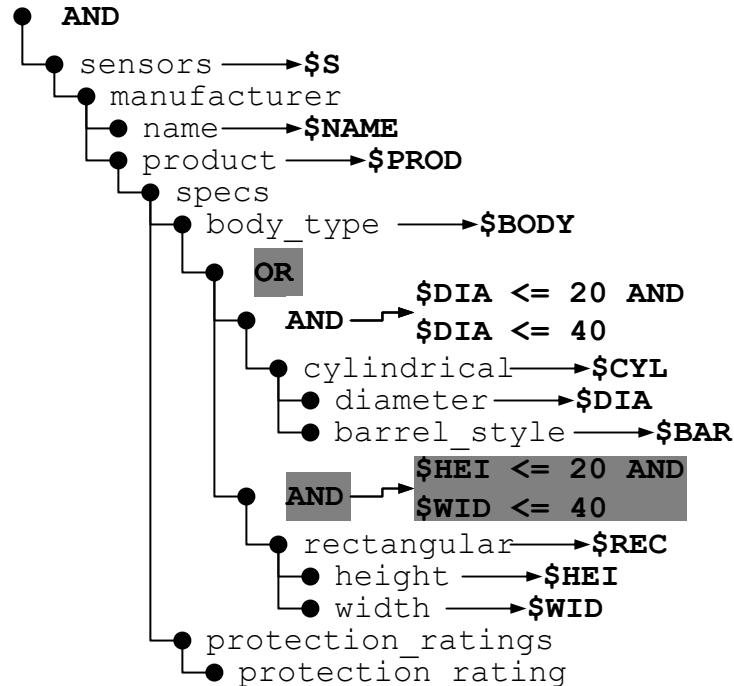
- Condition tree
- Schema-driven generation

- Result tree

# Tree Query Language (TQL)



# Tree Query Language (TQL)



→

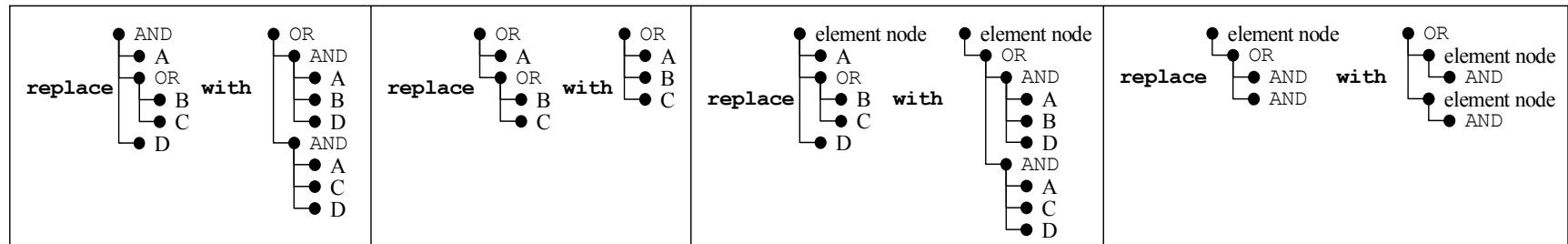
\$NAME	\$PROD	\$PART	\$BODY	\$CYL	\$DIA	\$BAR
Turck	product	A123	cylindrical	cylindrical	17	Smooth
	part_number			diameter		
				17		
\$NAME	\$PROD	\$PART	\$BODY			
Turck	product	B123	rectangular			
	part_number					
\$REC	\$HEI	\$WID				
	rectangular	10	30			
	height					
	10					

→

# TQL Semantics

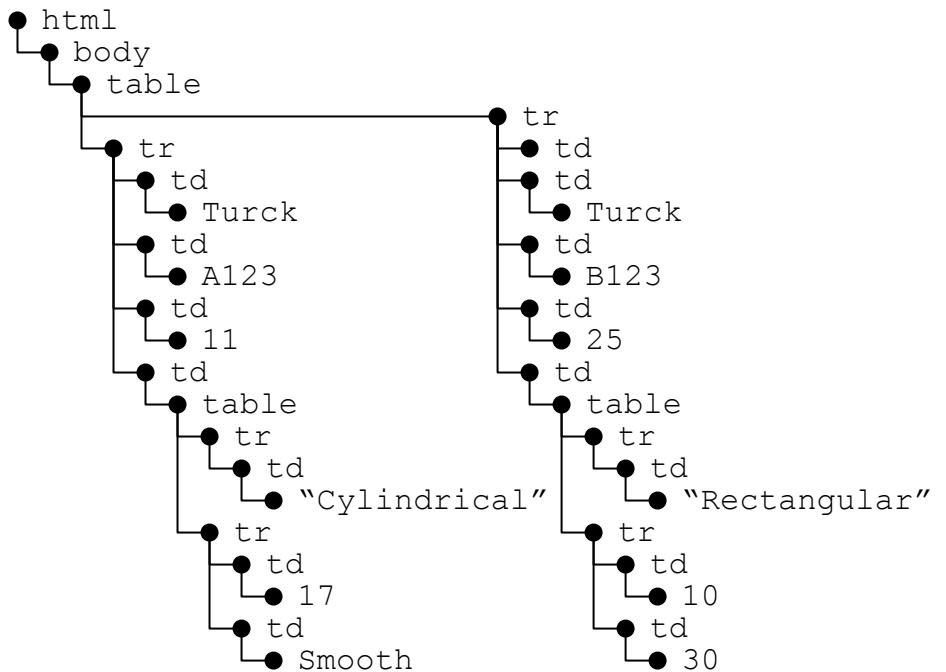
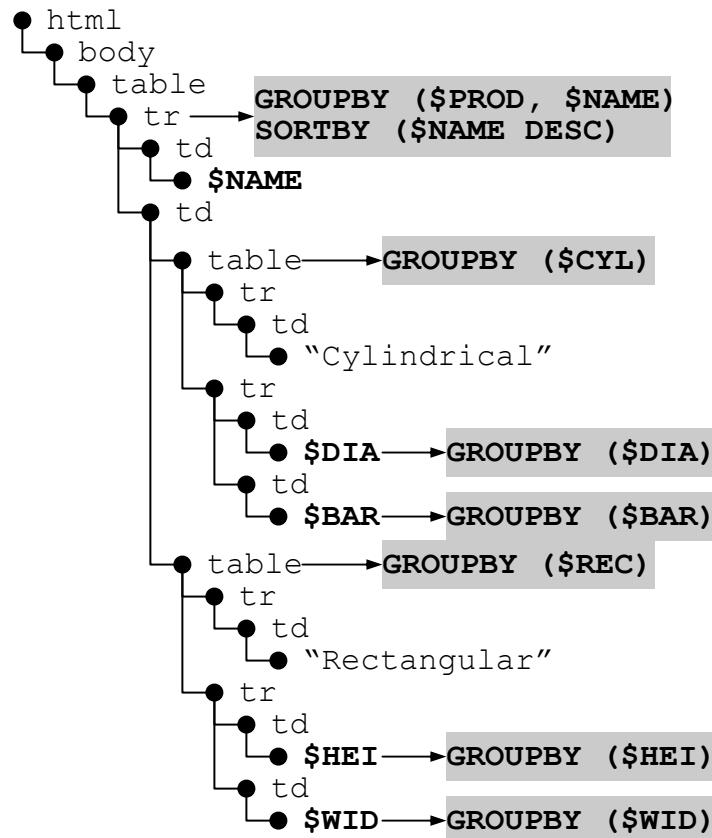
## Condition Tree

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



# TQL Semantics

## Result Tree



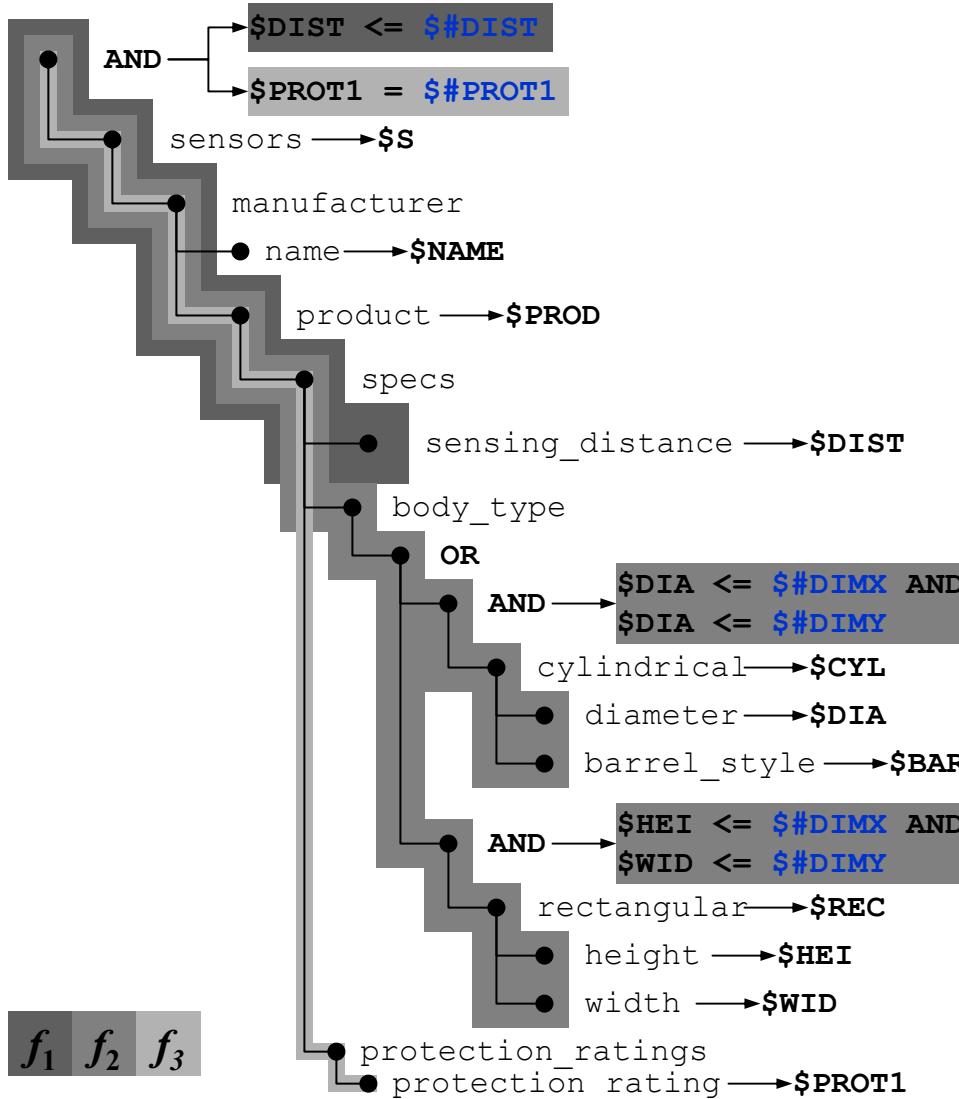
Name	Part Number	Image	Sensing Distance	Cylindrical	
				Diameter mm	Barrel Style
Turck	A123		11.0	17	Smooth
	B123		25.0	10	30
Rectangular		Height mm		Width mm	

# Tree Query Language (TQL)

---

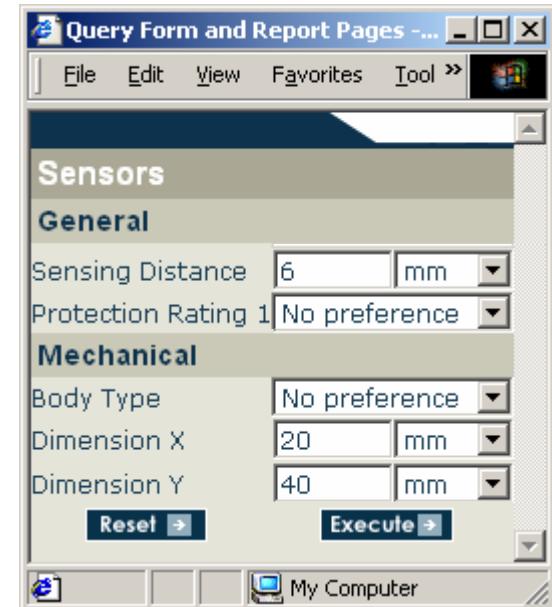
- Translated to XQuery
  - By QURSED Run-Time Engine
  - **TQL2XQuery Algorithm**
  - Syntax directed translation
  - Tree patterns in TQL to nested FOR-WHERE-RETURN expressions in XQuery

# Query Set Specification (QSS)

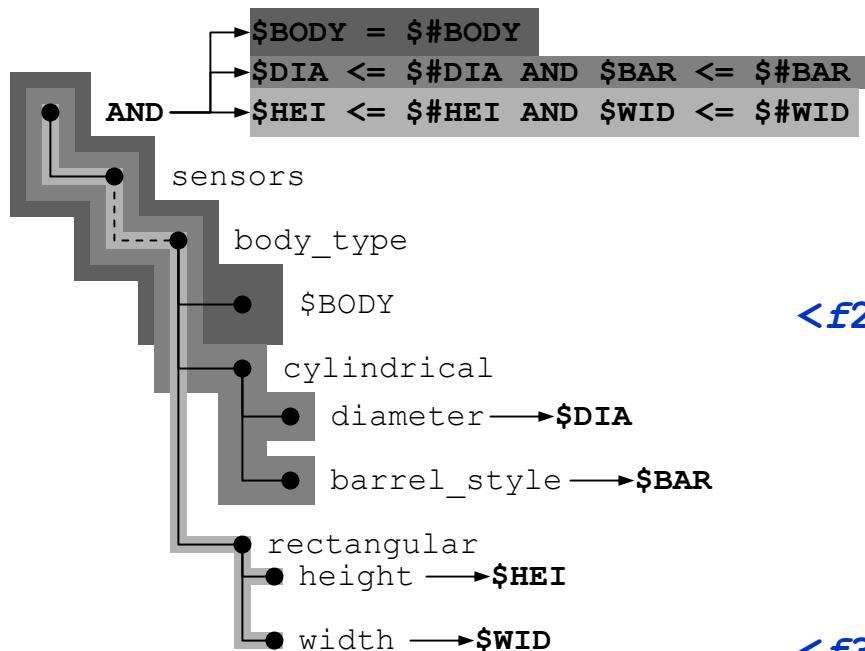


## Condition Tree Generator

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



# Dependencies



## Mechanical

Body Type	Cylindrical
Diameter	mm
Barrel Style	No preference

$\langle f2, \ $#BODY = \text{"Cylindrical"}, \ f1 \rangle$

## Mechanical

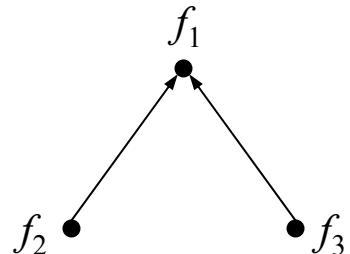
Body Type	Rectangular
Height	mm
Width	mm

$\langle f3, \ $#BODY = \text{"Rectangular"}, \ f1 \rangle$

$f_1 \ f_2 \ f_3$

# Dependencies

---

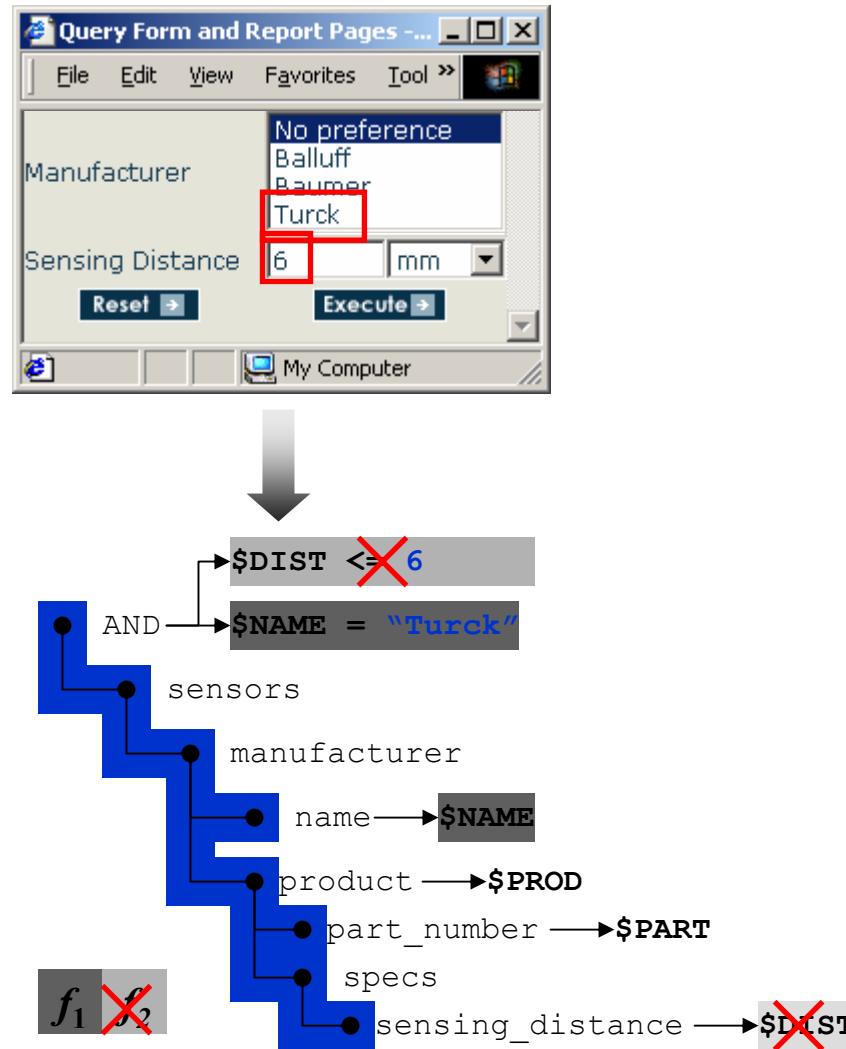


`<f2, #BODY = "cylindrical", {f1}>`

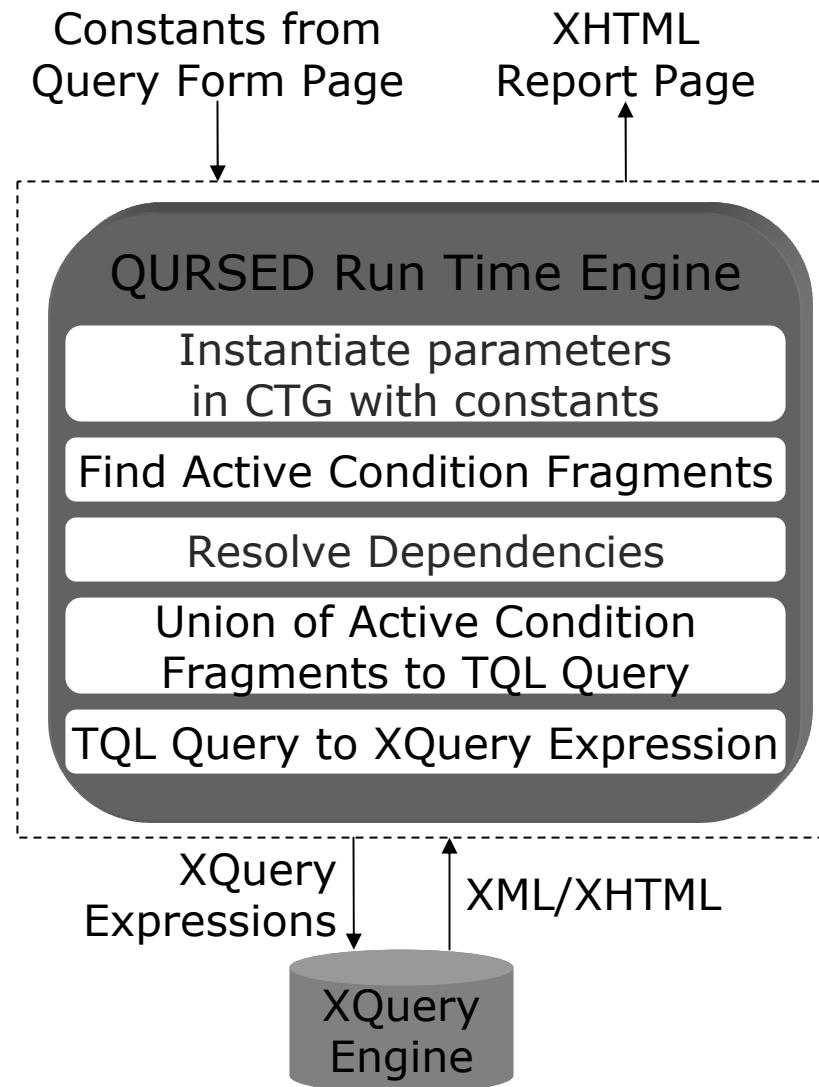
`<f3, #BODY = "rectangular", {f1}>`

- Dependencies Graph
- Resolution algorithm based on topological sort

# Run-time: QSS to TQL Queries

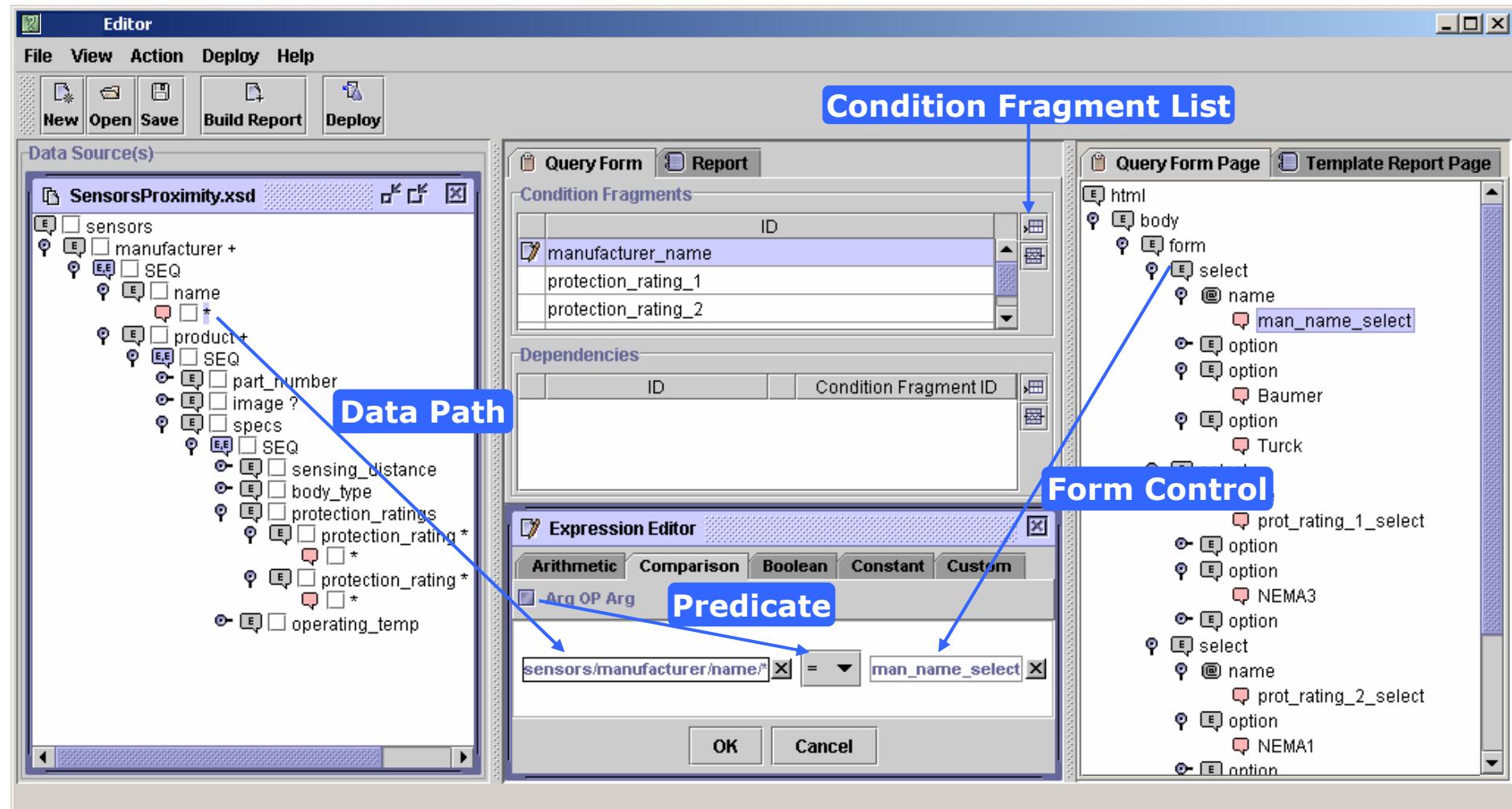


# Run-time: QSS to TQL Queries



# QURSED Editor

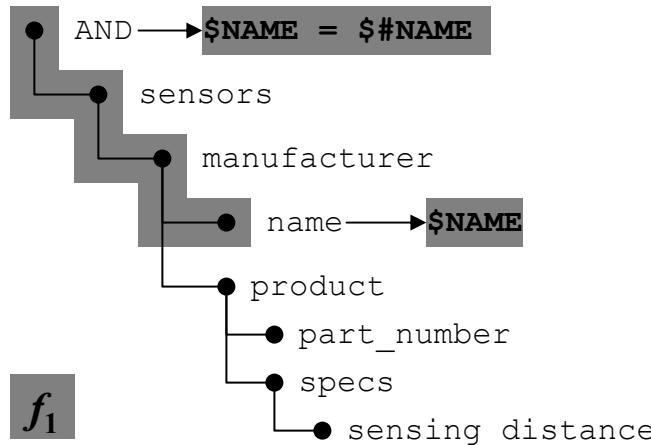
## Building Query/Visual Association



# QURSED Editor

## From visual actions to QSS

```
sensors/manufacturer/*  
= man_name_select
```

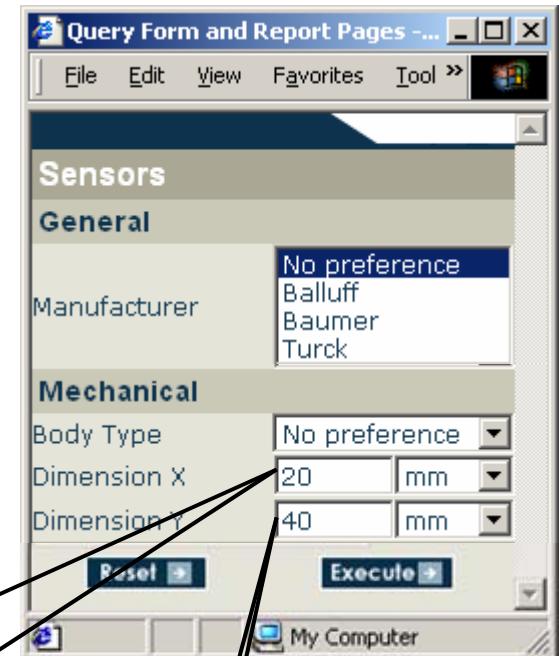
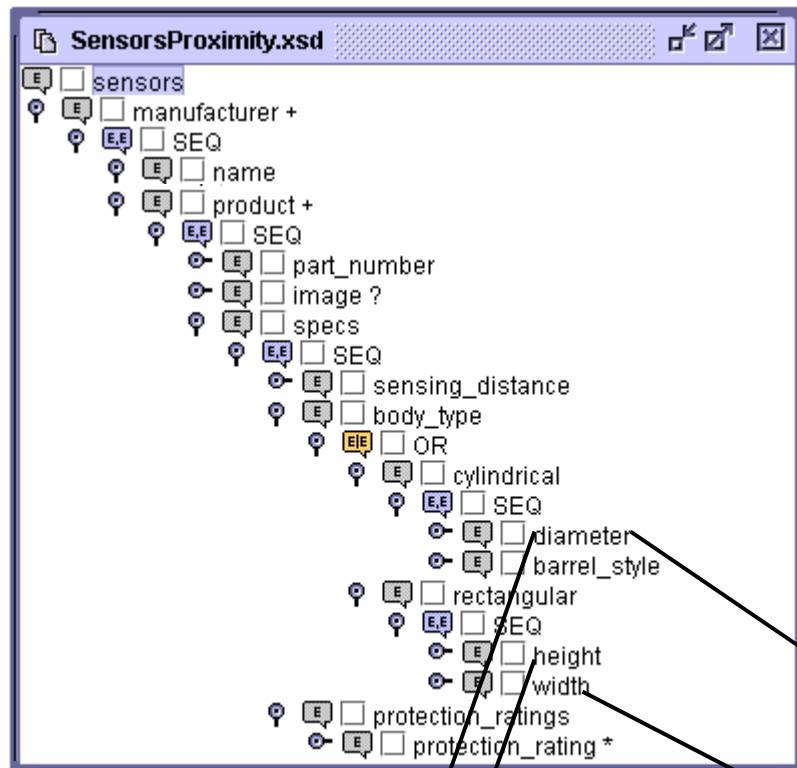


Choice of schema element e means

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

# QURSED Editor

## Disjunction



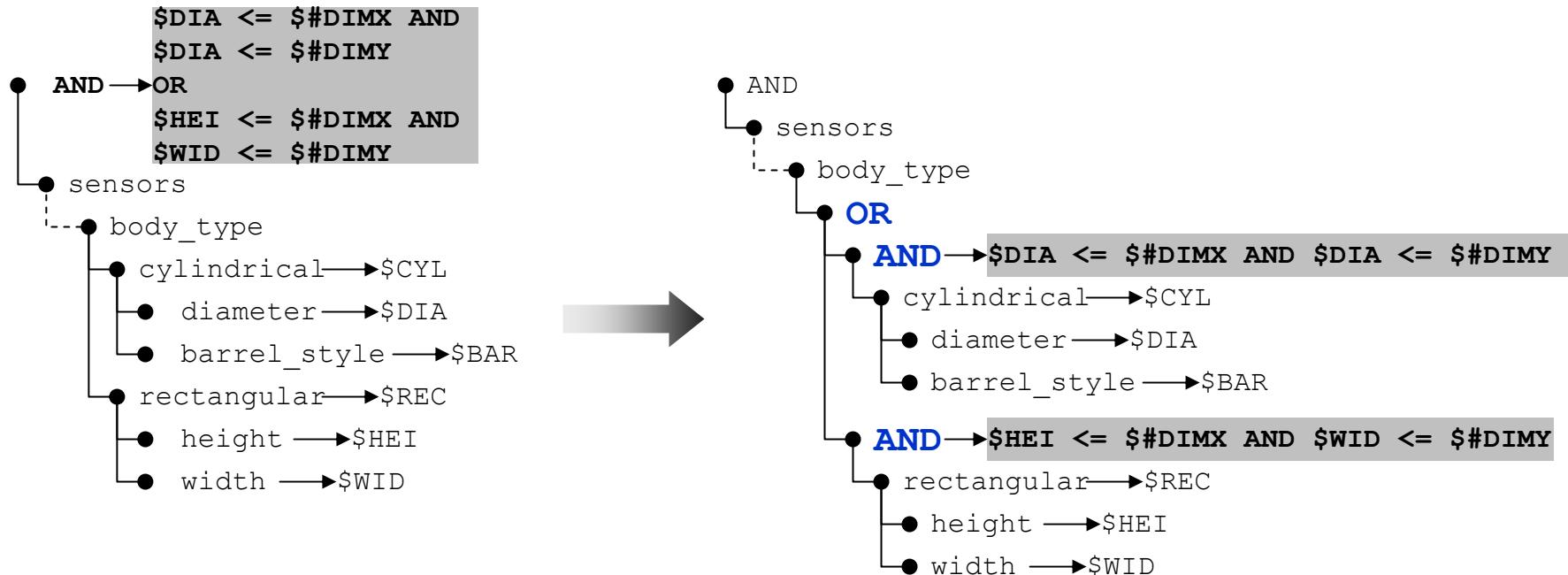
$$(\$DIA \leq \$\#DIMX \text{ AND } \$DIA \leq \$\#DIMY)$$

~~OR~~

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

# QURSED Editor

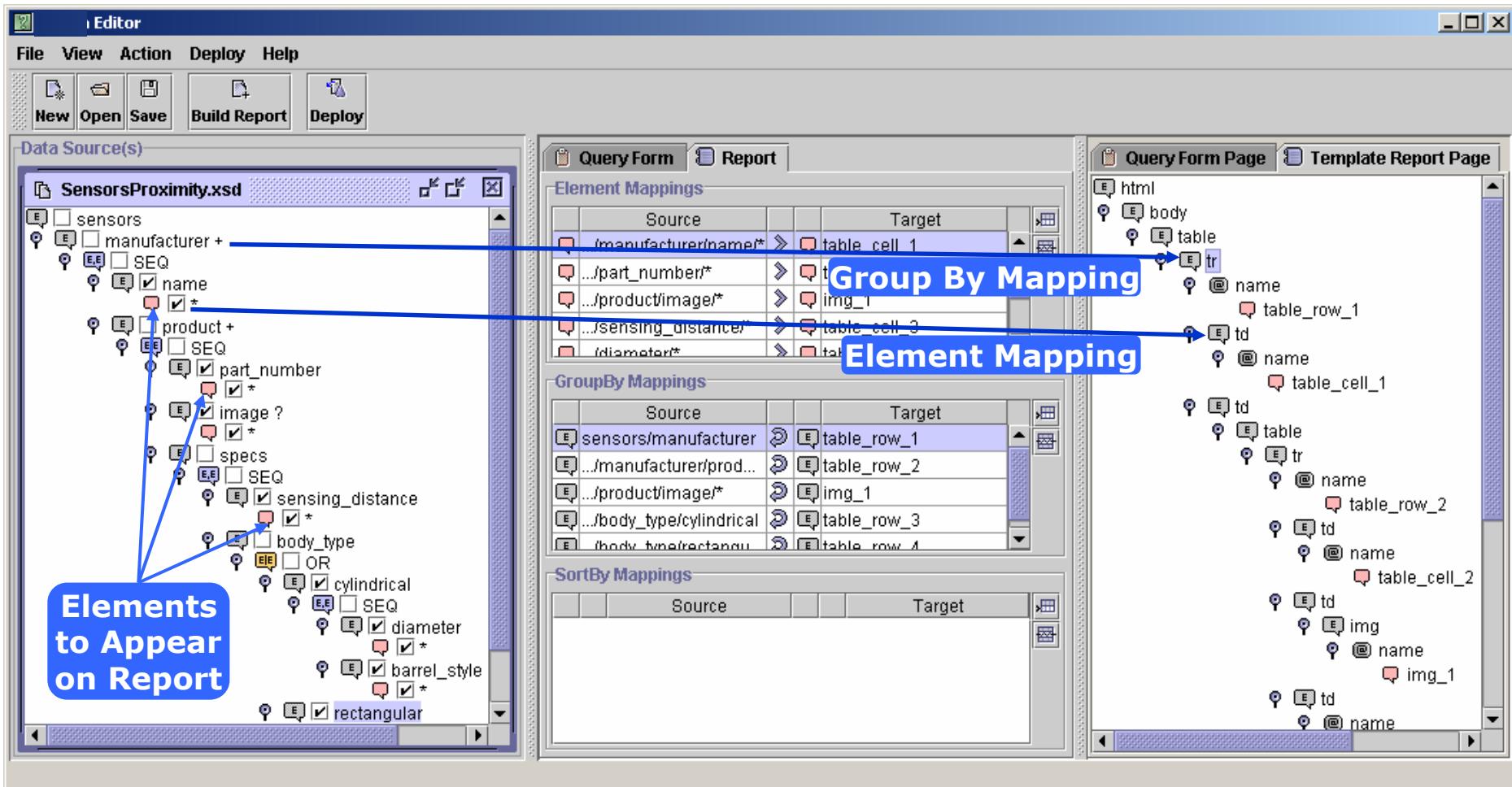
## Disjunction



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

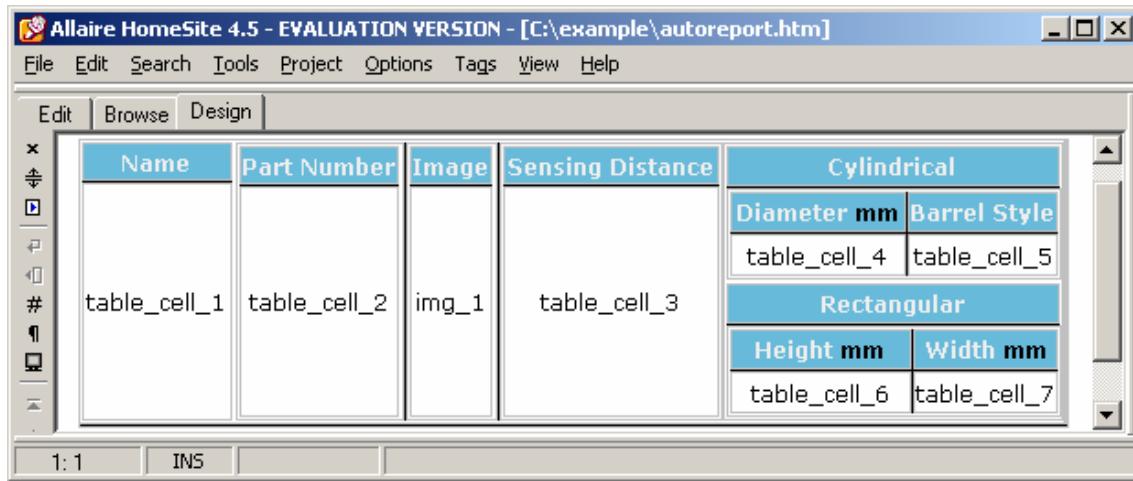
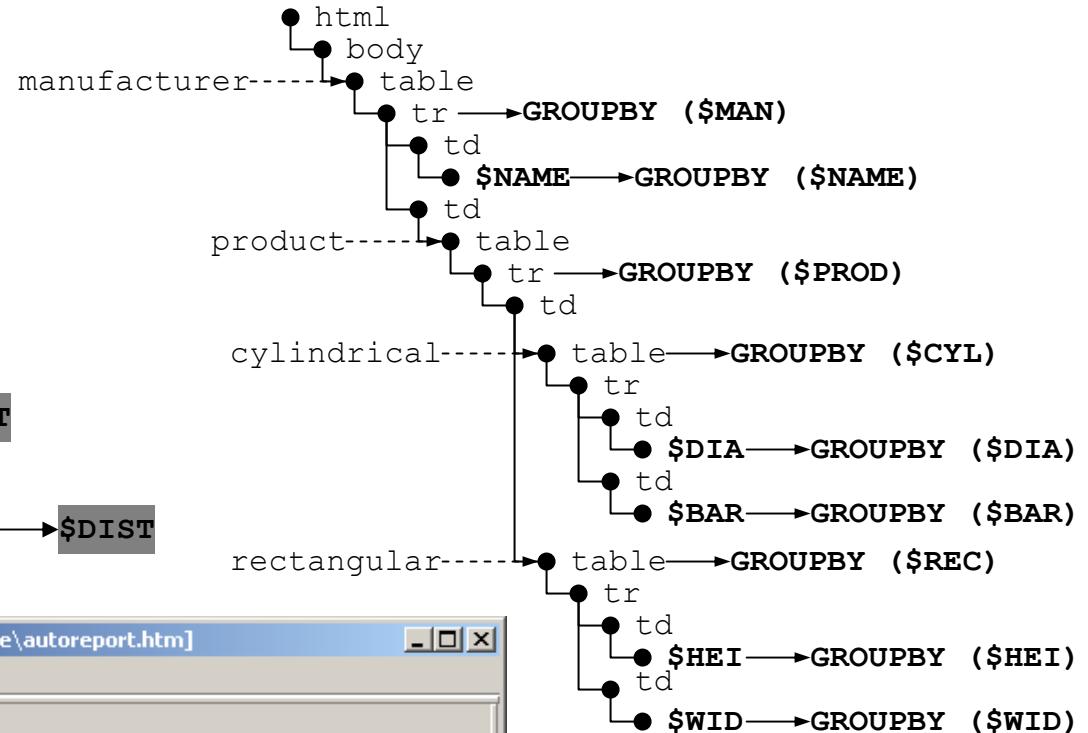
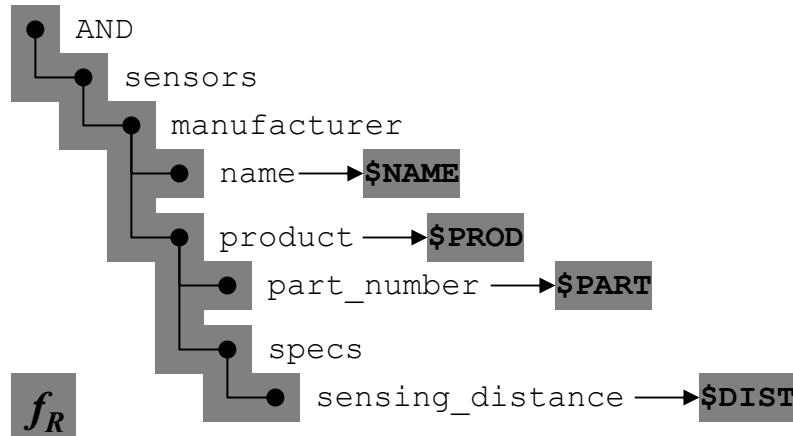
# QURSED Editor

# Building Reports



# 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

---

?