QURSED: Querying and Reporting Semistructured Data

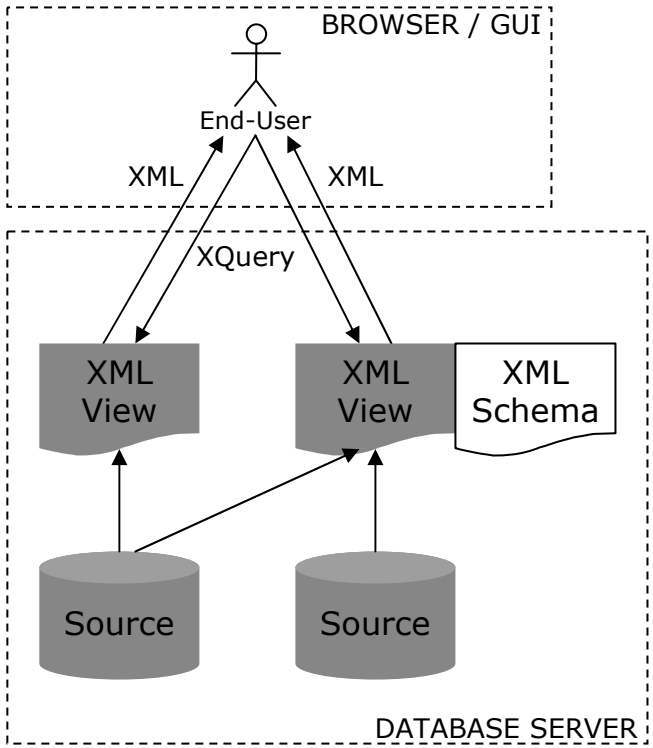
Yannis Papakonstantinou
Michalis Petropoulos
UNIVERSITY OF CALIFORNIA, SAN DIEGO
Vasilis Vassalos
NEW YORK UNIVERSITY

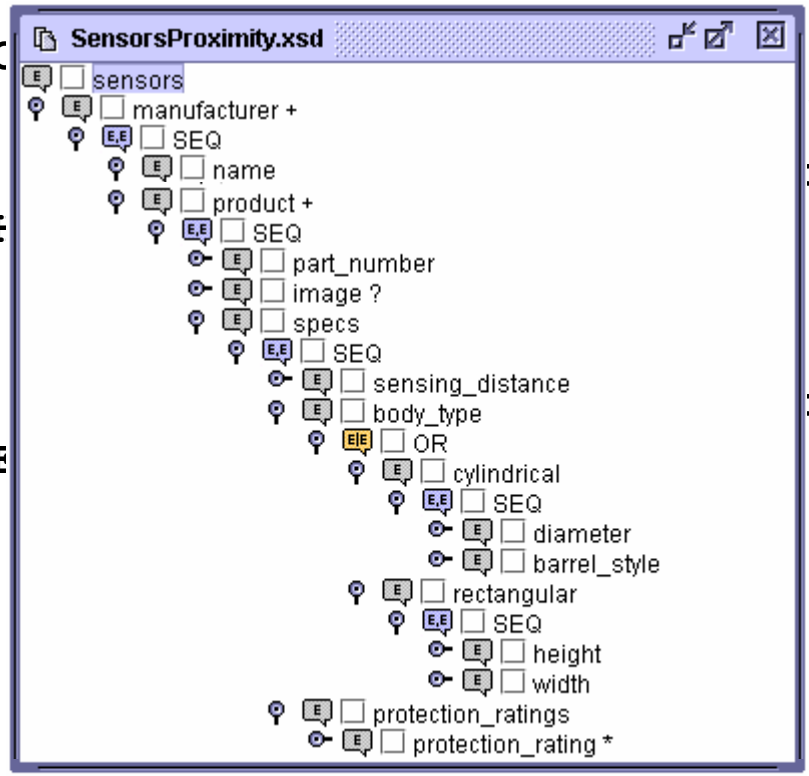
June 2002

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

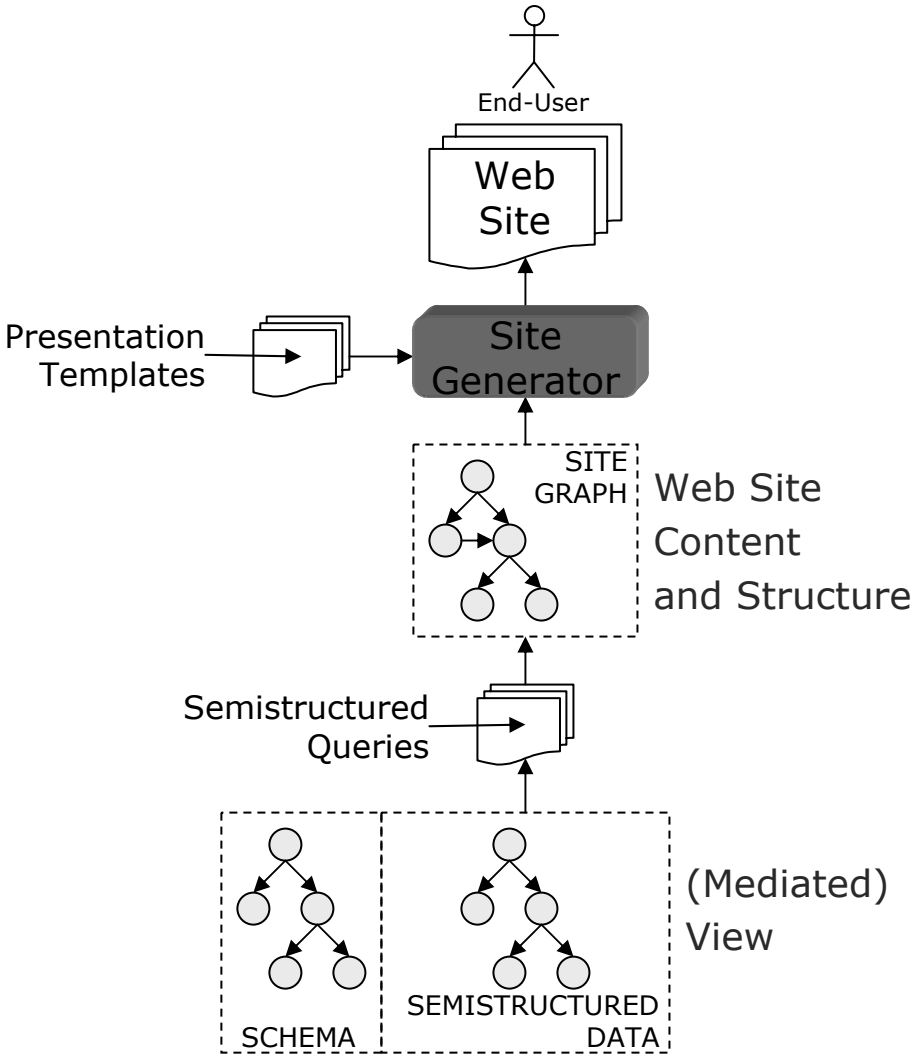
Exporting DBMSs on the Web





- 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

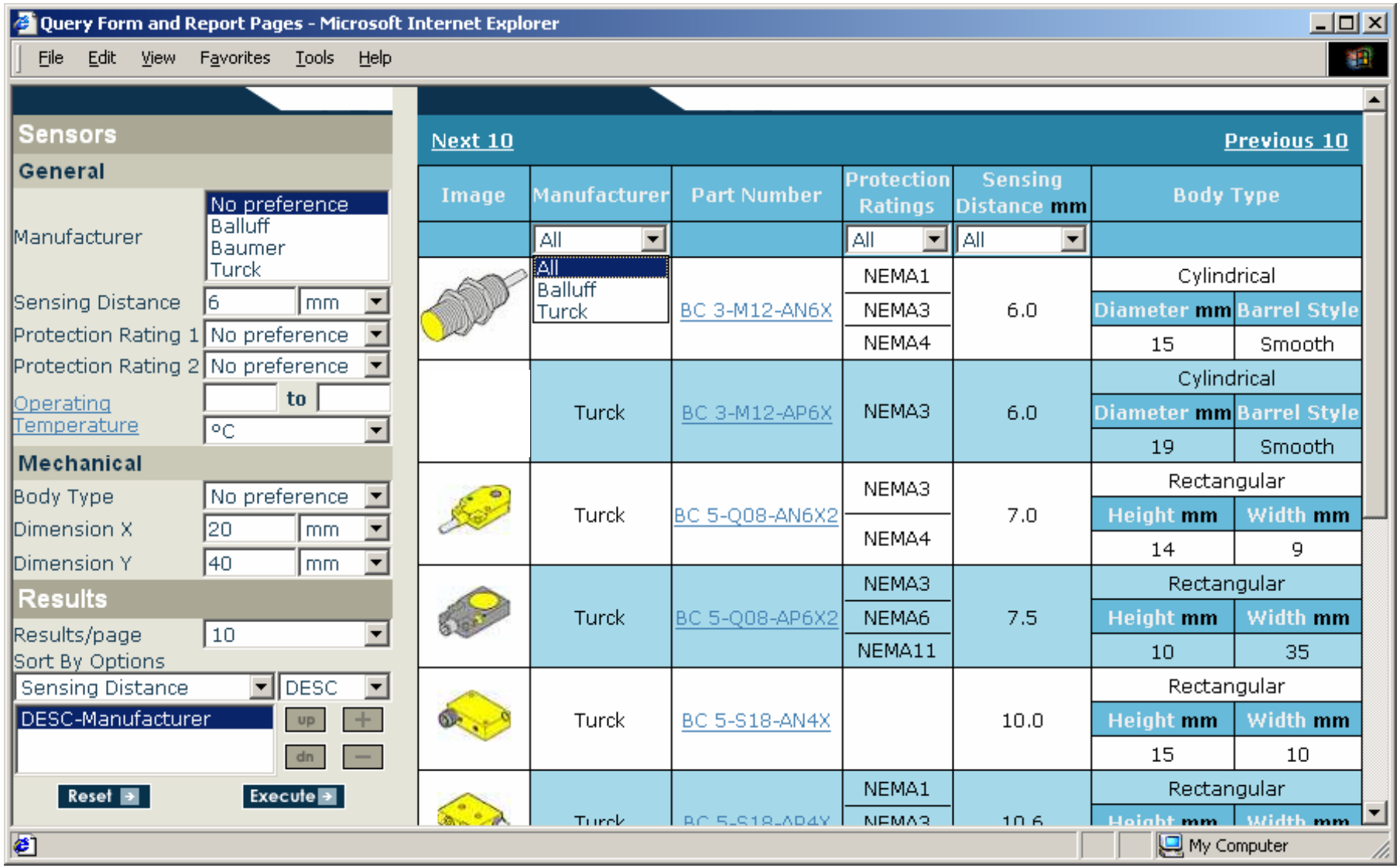
- Declarative
- Separation of content, structure and presentation

Query Forms and Reports

Requirements

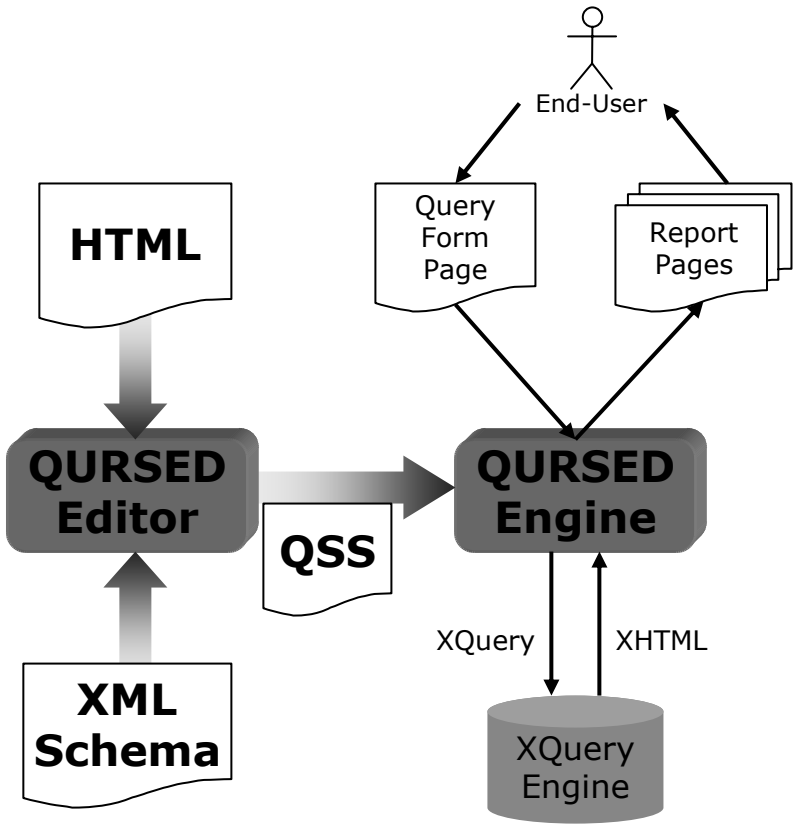
- Handle semistructureness
  - Powerful query forms and reports
- Be declarative
  - Separate logic from presentation
- Encode compactly a large number of queries
  - Compared to a set of query templates
- Visual interface for the developer
  - Programming should NOT be a requirement

Query Forms and Reports



Query Forms and Reports

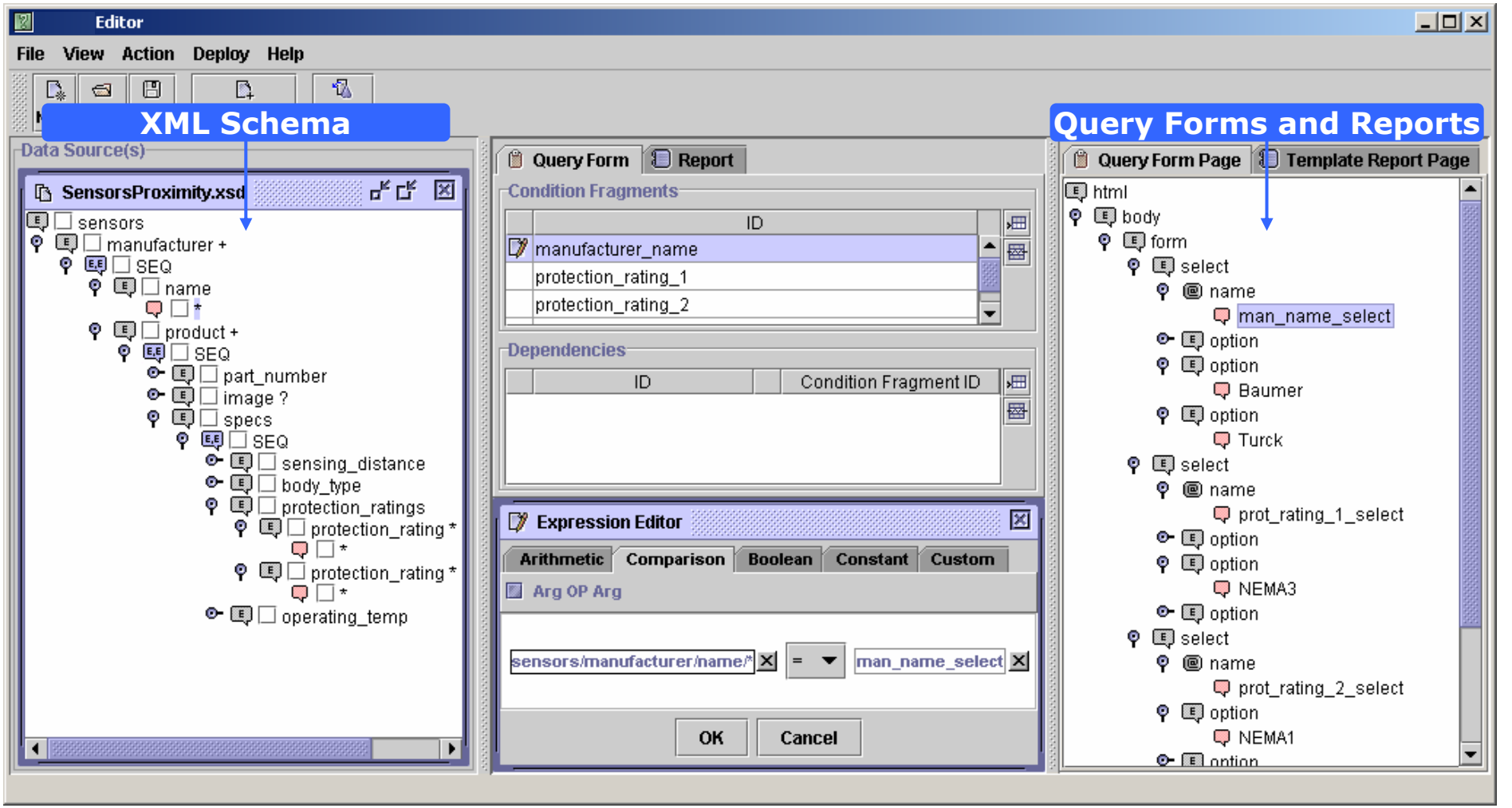
QURSED Approach



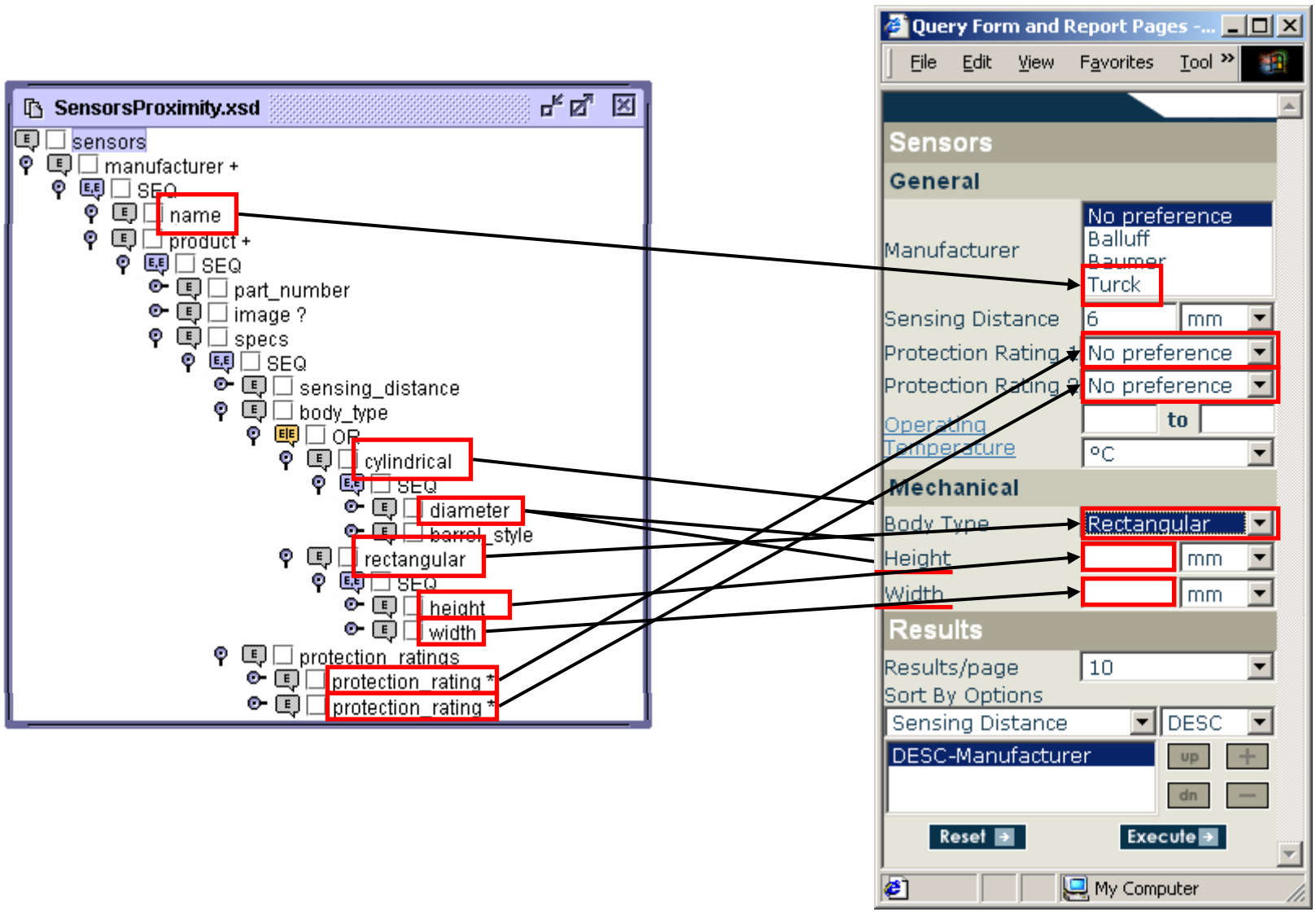
- XML Schema-driven
- Declarative!
  - Separation of content & presentation
- Editor
  - Visual actions to declarative specifications
  - Automatic construction of report pages
- Query Set Specification (QSS)
  - Large set of parameterized queries
  - Compact representation
- Engine
  - Automatic query formulation
  - Direct result construction

Query Forms and Reports

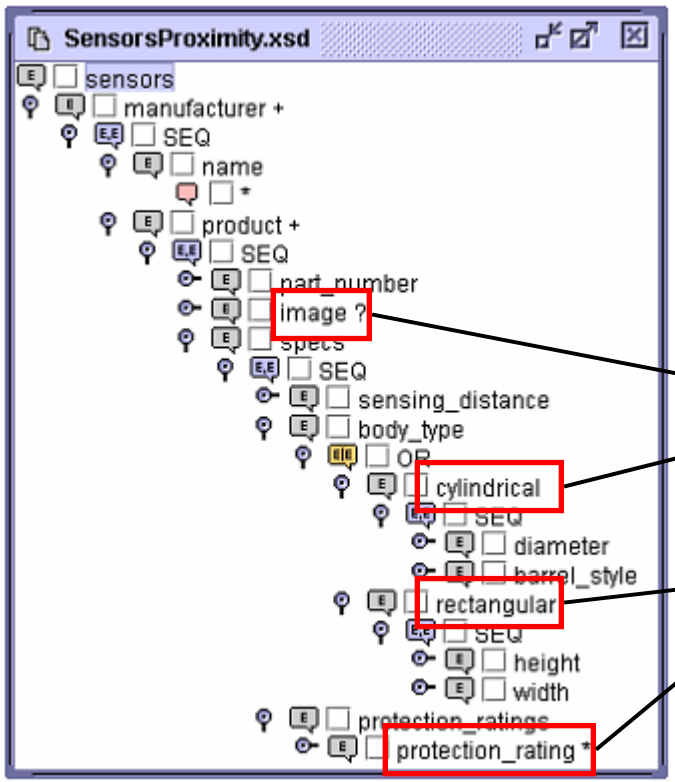
QURSED Editor

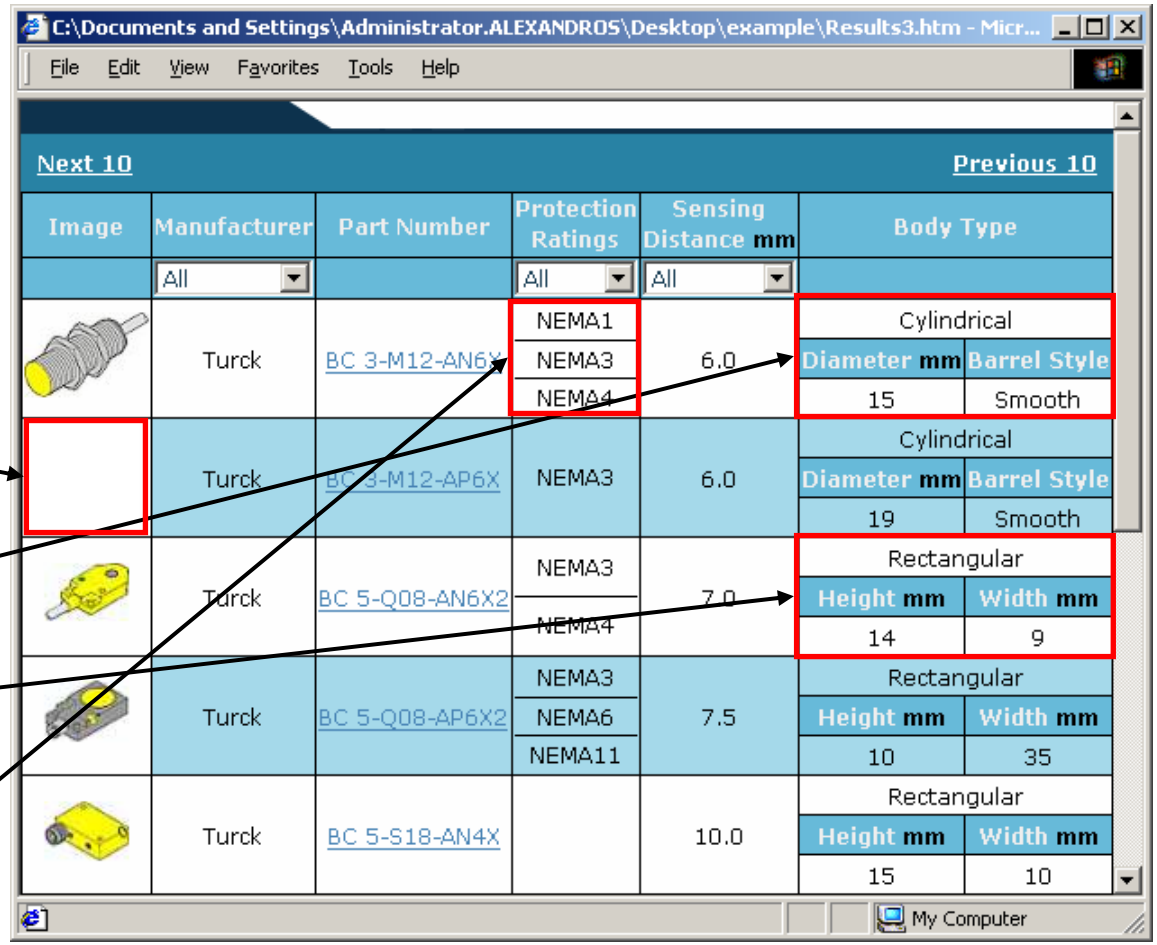


Developing Query Forms from the XML Schema

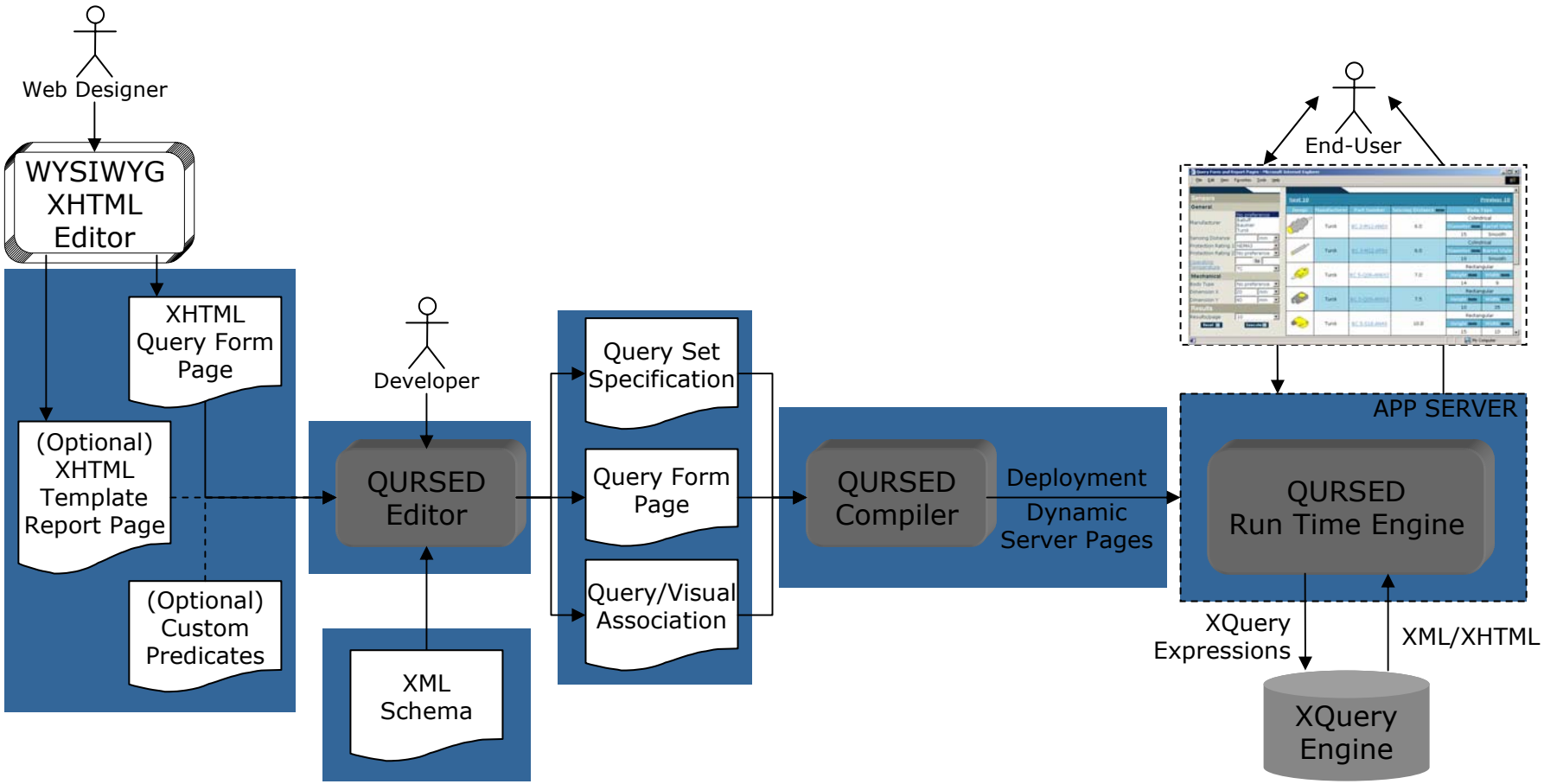


Developing Reports from the XML Schema

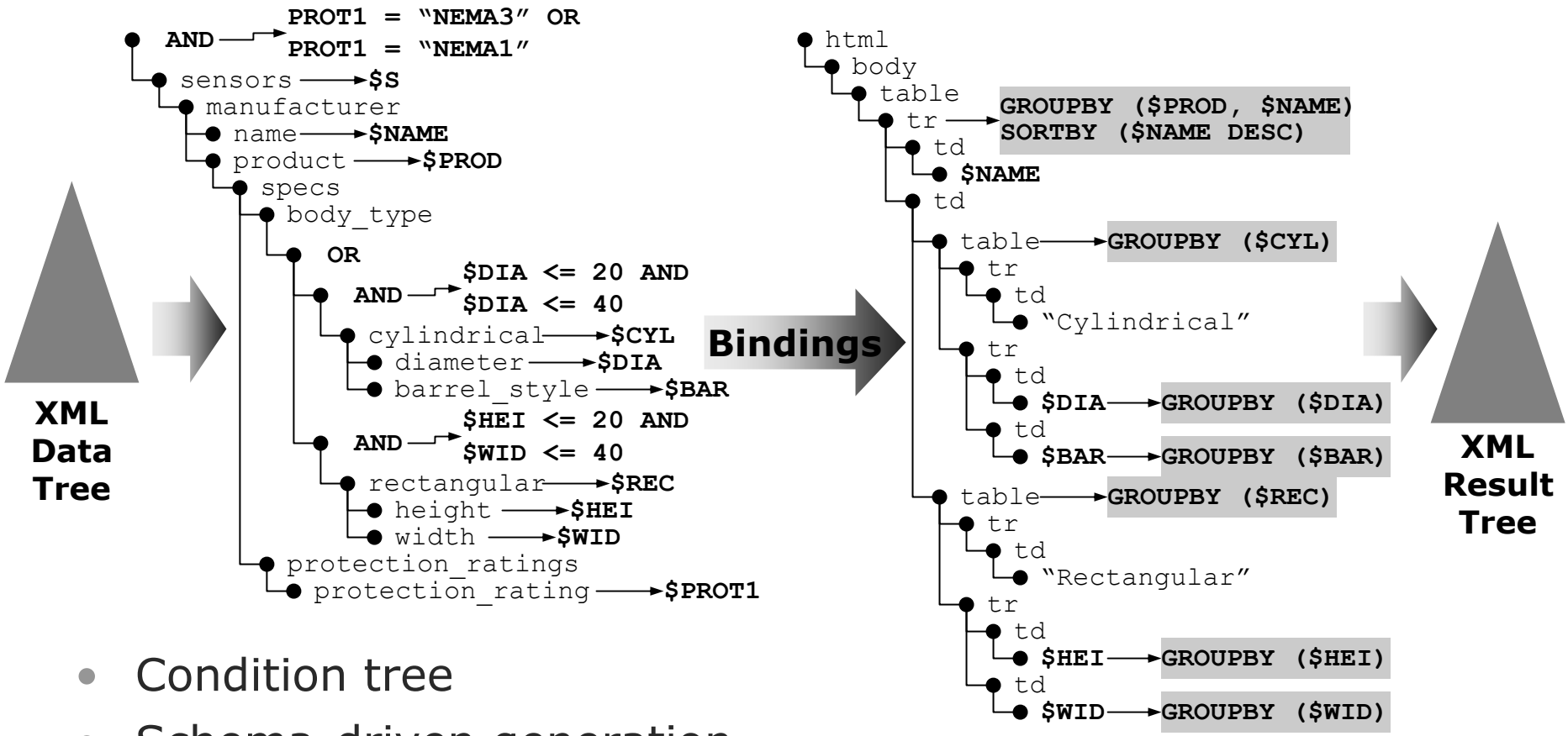




QURSED System Architecture

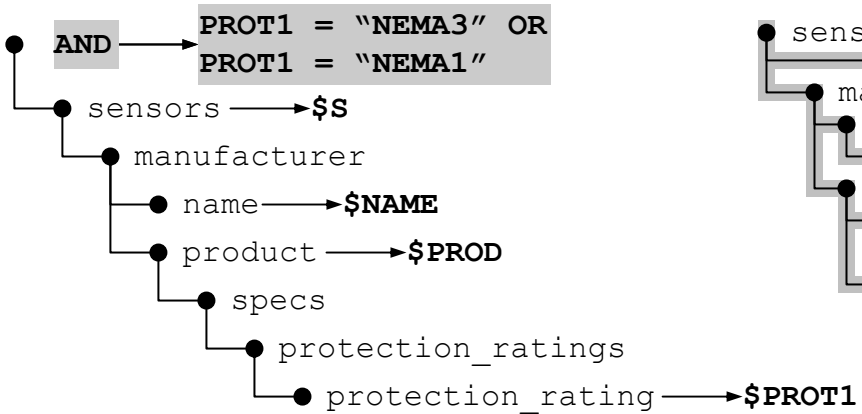


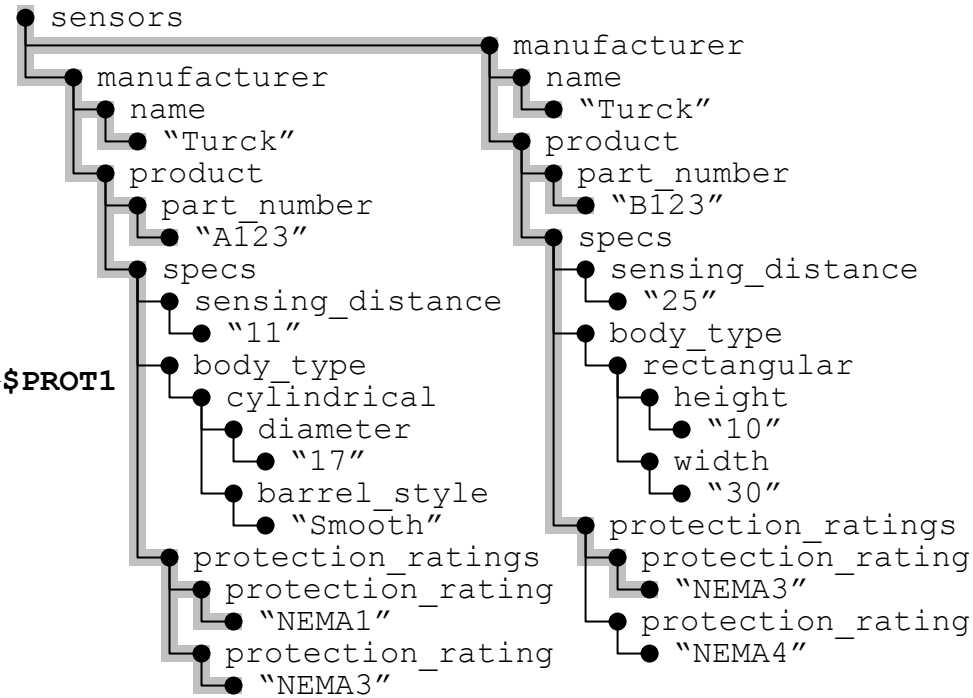
Tree Query Language (TQL)



- Condition tree
- Schema-driven generation
- Result tree

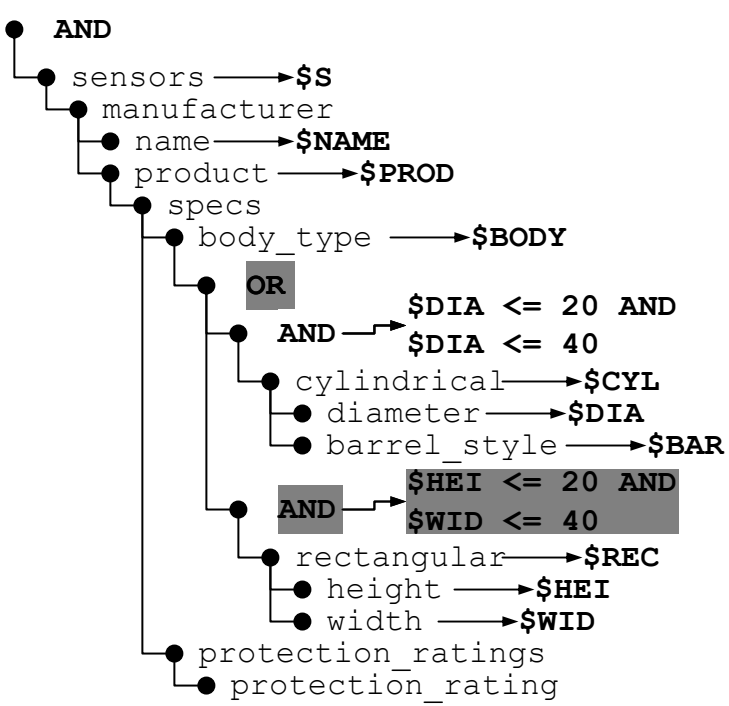
Tree Query Language (TQL)

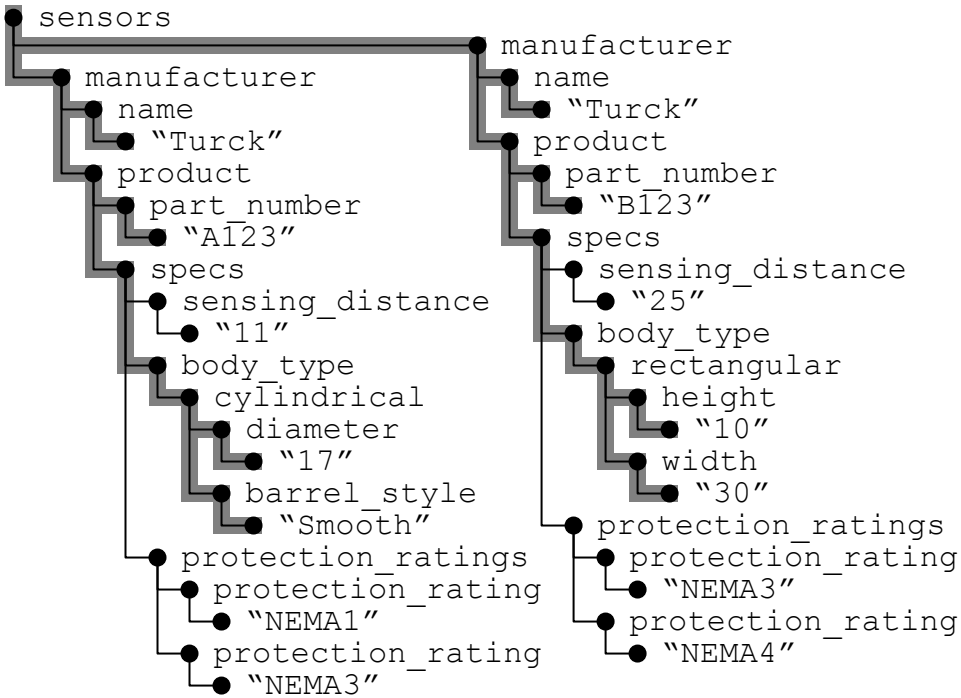




$NAME	$PROD	$PART	$PROT1
Turck	product / part_number / "A123"	A123	NEMA1
Turck	product / part_number / "A123"	A123	NEMA3
Turck	product / part_number / "B123"	B123	NEMA3

Tree Query Language (TQL)





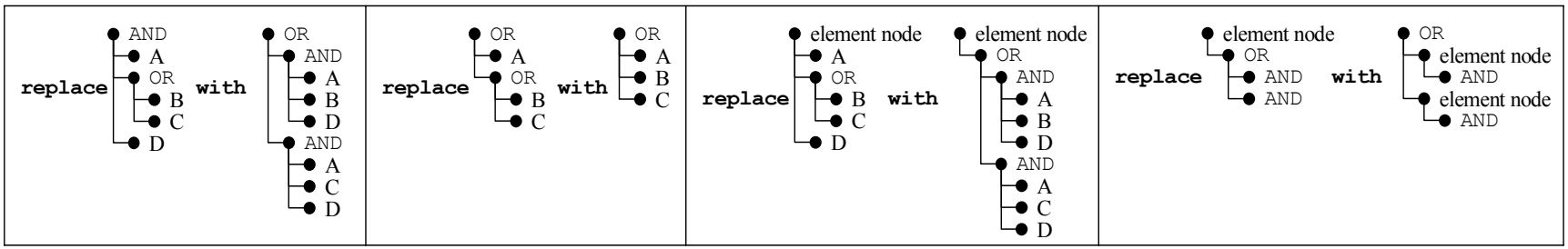
$NAME	$PROD	$PART	$BODY	$CYL	$DIA	$BAR
Turck	product / part_number / "A123"	A123	cylindrical	cylindrical / diameter / "17"	17	Smooth

$NAME	$PROD	$PART	$BODY	$REC	$HEI	$WID
Turck	product / part_number / "B123"	B123	rectangular	rectangular / height / "10"	10	30

TQL Semantics

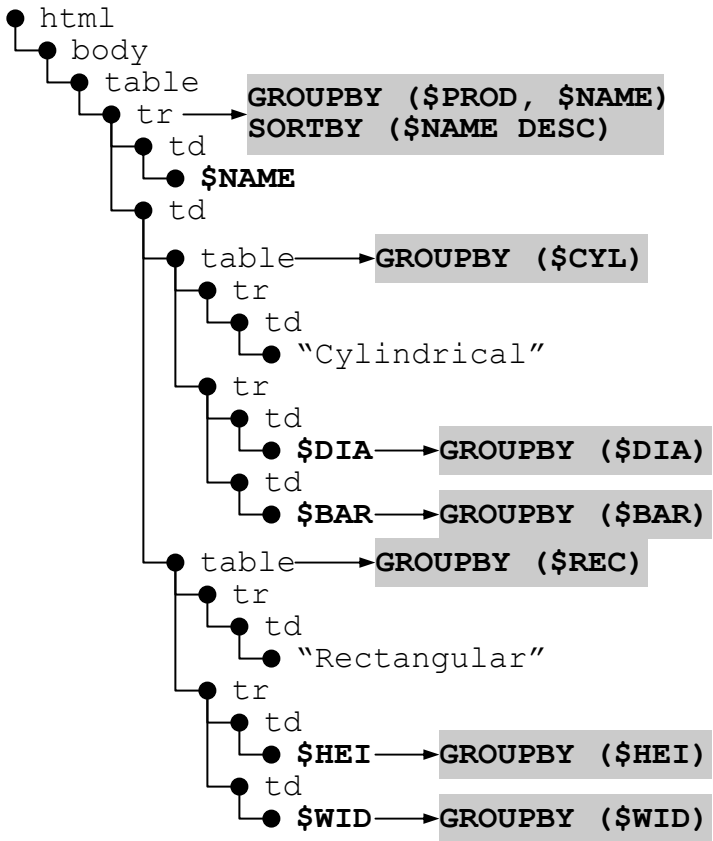
Condition Tree

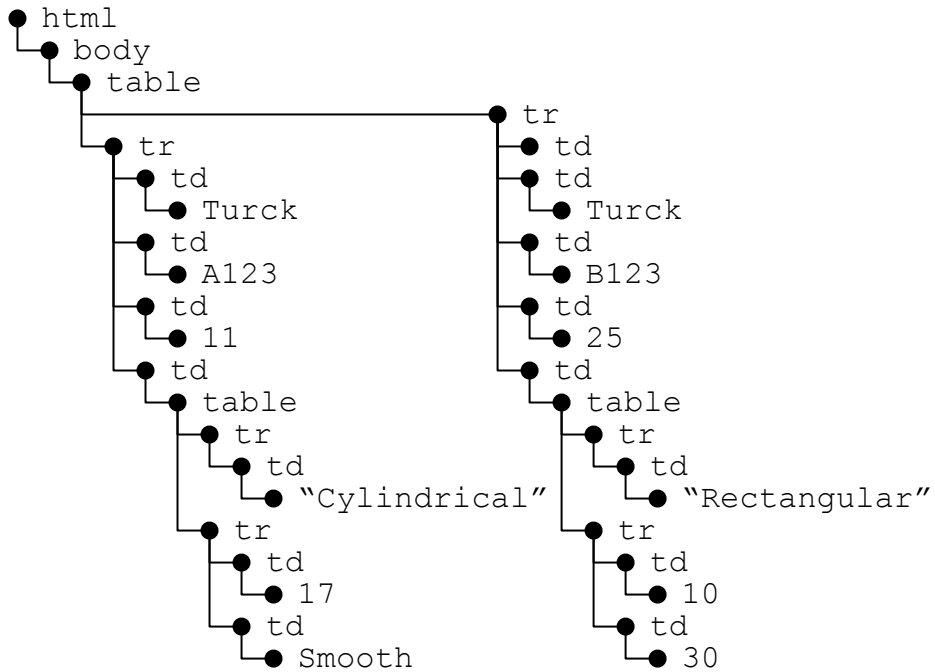
- Conjunctive Condition Trees
  - **OR-Removal** Algorithm
  - Transformation Rules



TQL Semantics

Result Tree



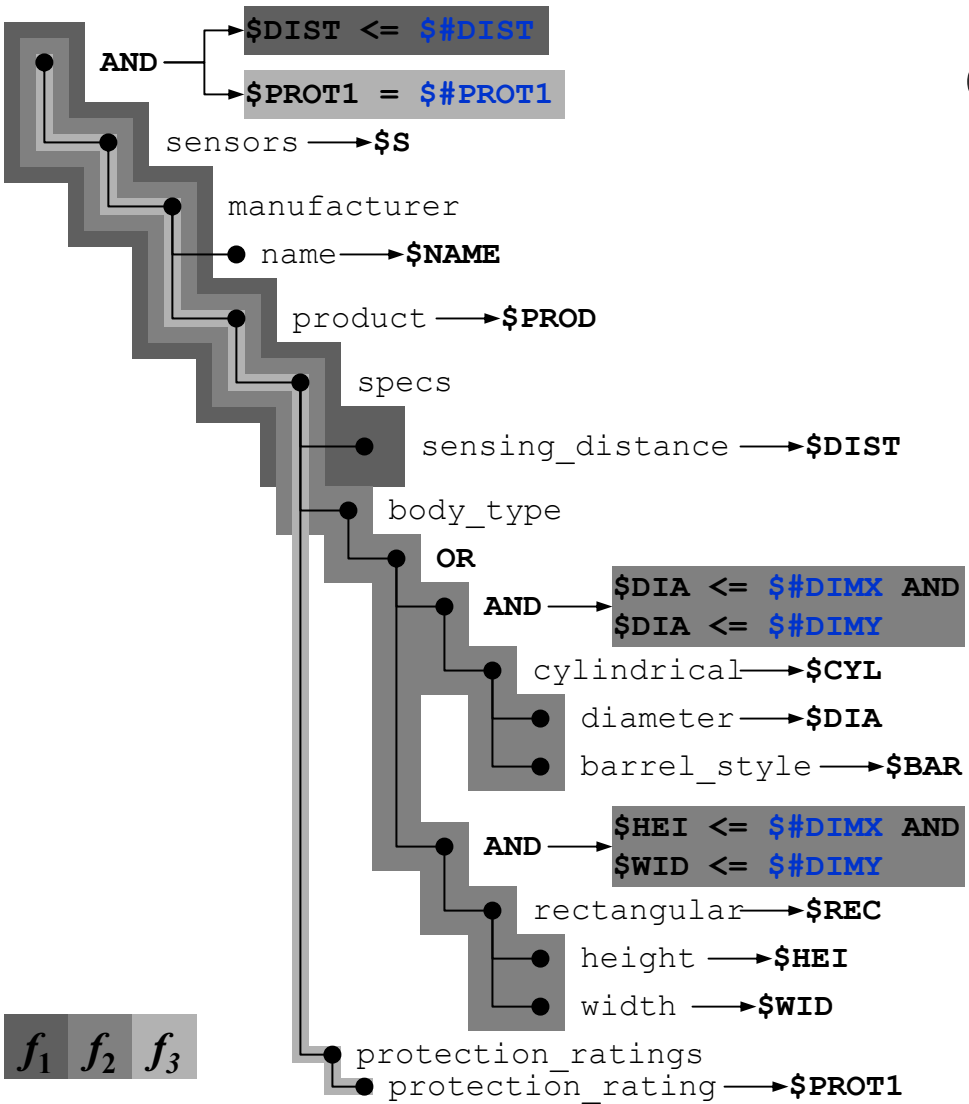


Name	Part Number	Image	Sensing Distance	Cylindrical	
				Diameter mm	Barrel Style
Turck	A123		11.0	17	Smooth
	Part Number	Image	Sensing Distance	Rectangular	
				Height mm	Width mm
	B123		25.0	10	30

Tree Query Language (TQL)

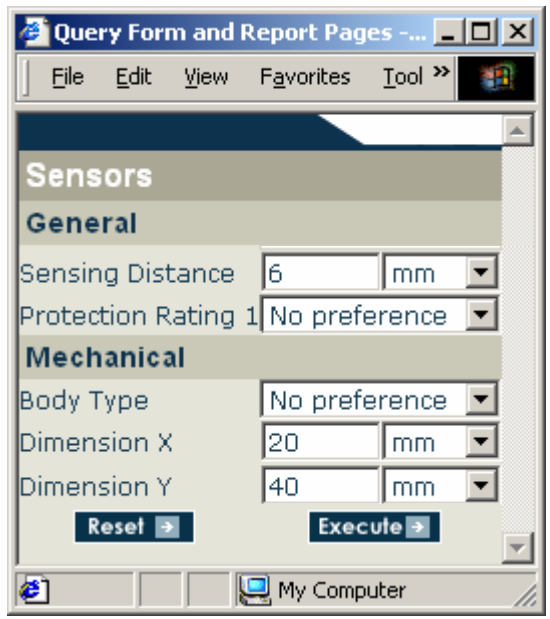
- Translated to XQuery
  - By QURSED Run-Time Engine
  - **TQL2XQuery Algorithm**
  - Syntax directed translation
  - Tree patterns in TQL to nested FOR-WHERE-RETURN expressions in XQuery

Query Set Specification (QSS)

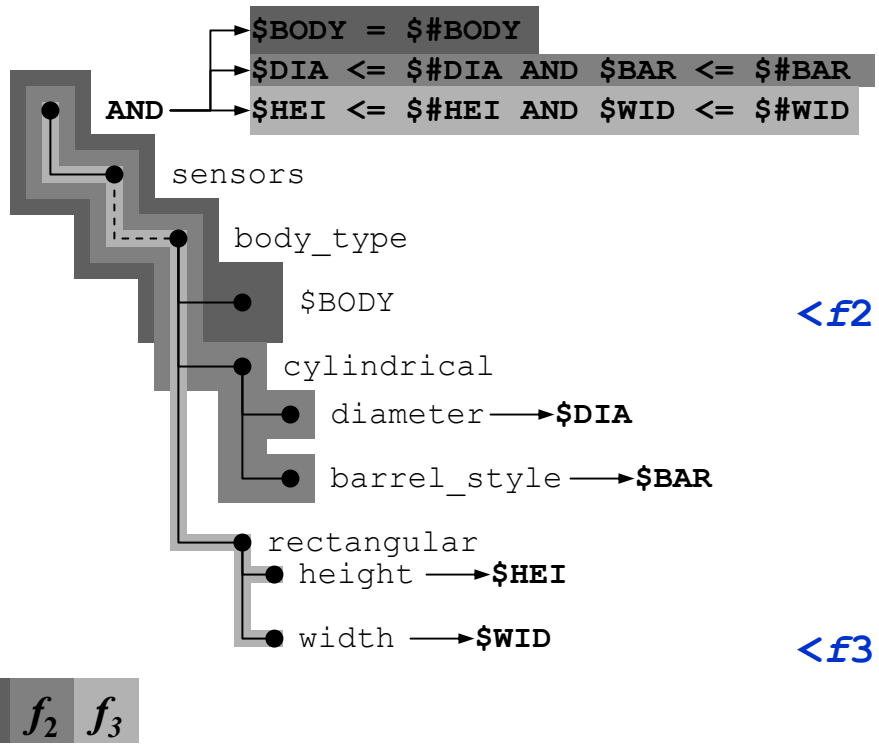


Condition Tree Generator

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



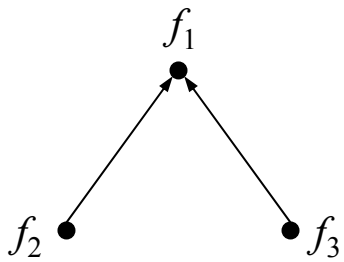
Dependencies



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

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

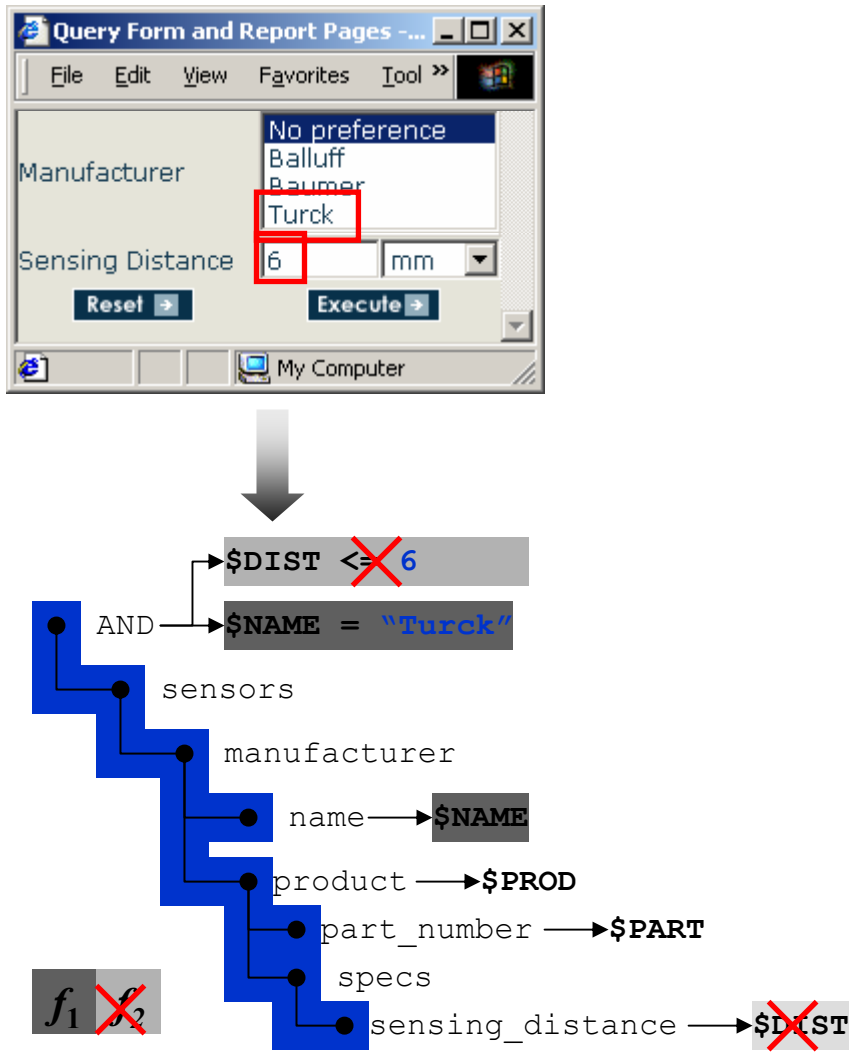
Dependencies



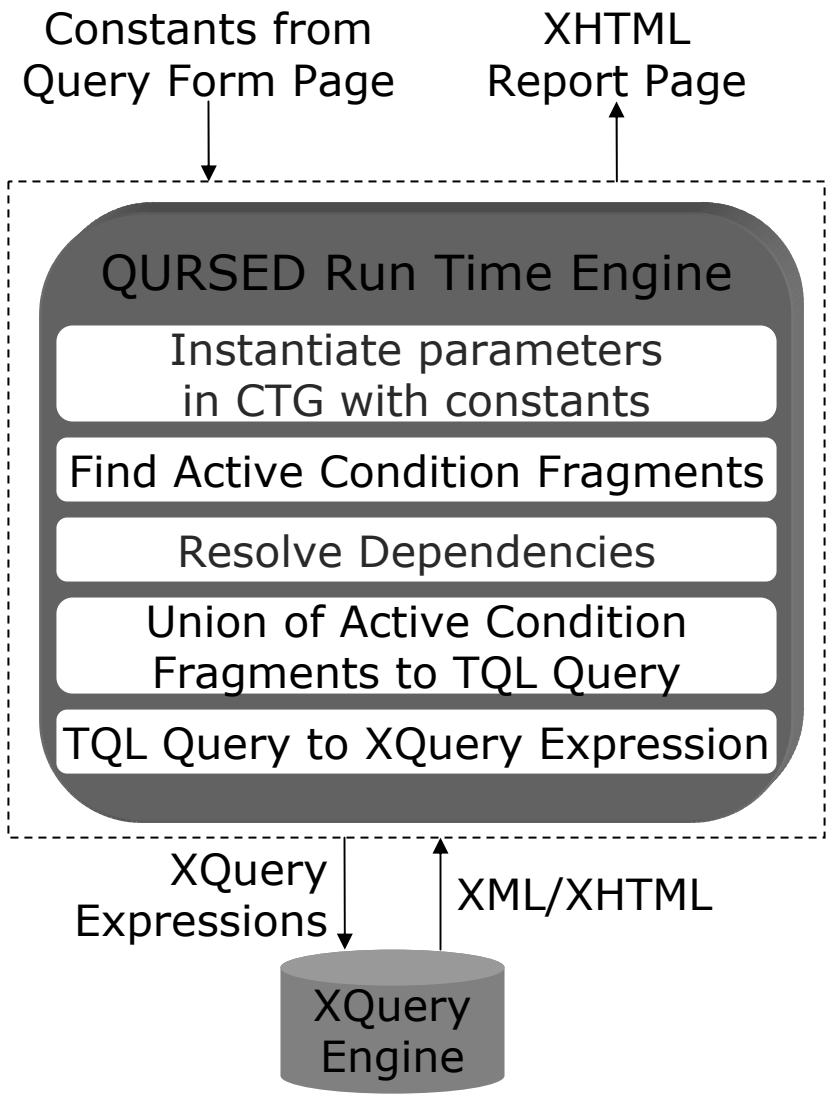

```
<f2, $#BODY = "cylindrical", {f1}>
<f3, $#BODY = "rectangular", {f1}>
```

- Dependencies Graph
- Resolution algorithm based on topological sort

Run-time: QSS to TQL Queries

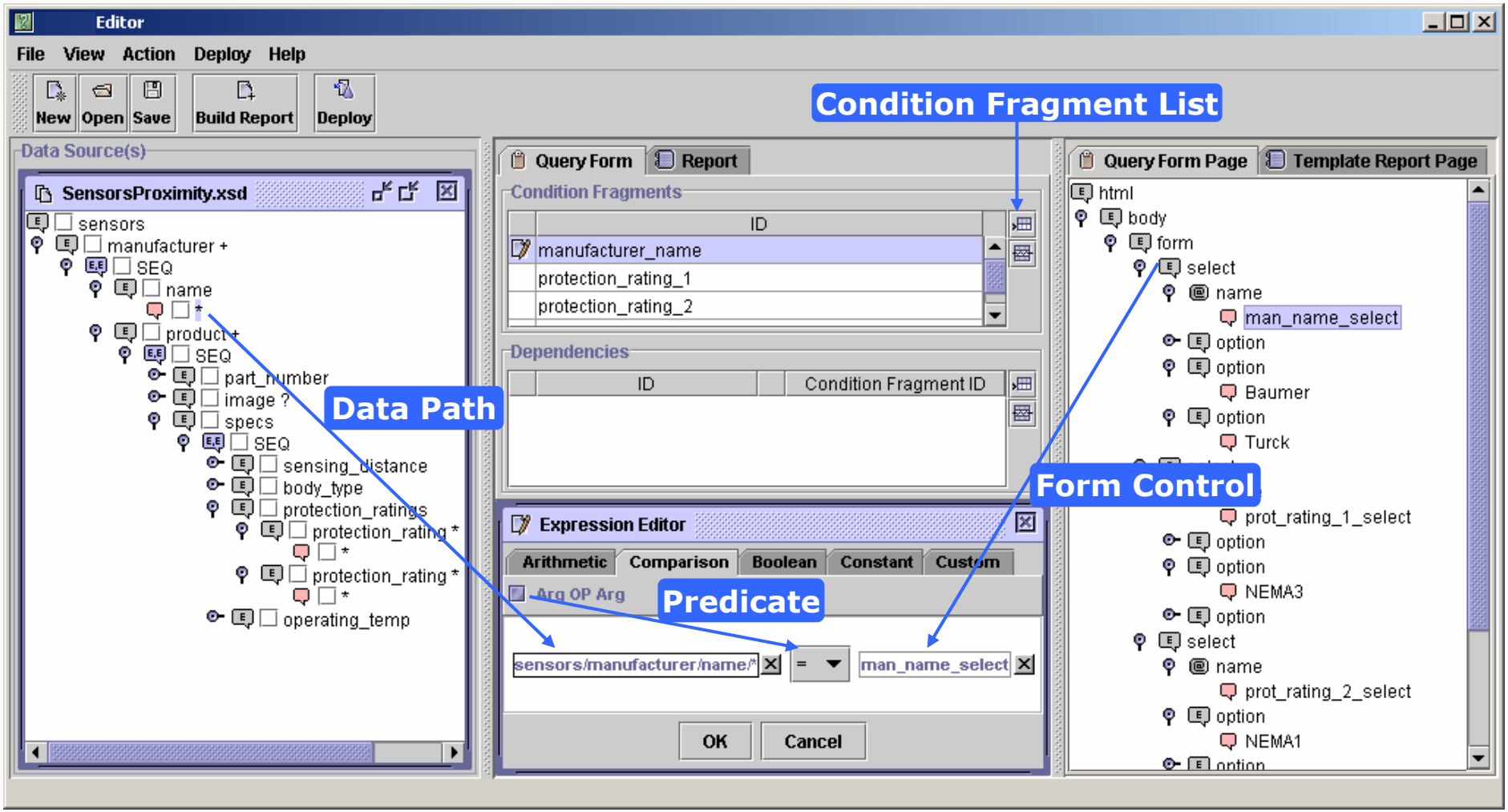


Run-time: QSS to TQL Queries



QURSED Editor

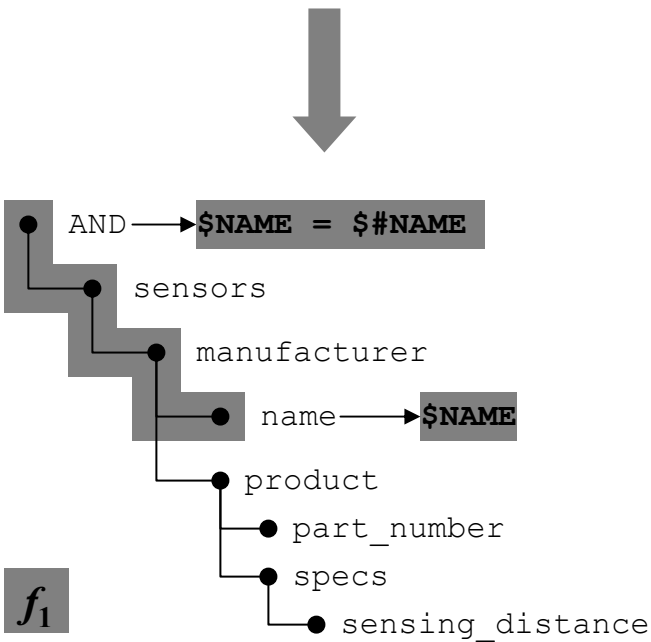
Building Query/Visual Association



QURSED Editor

From visual actions to QSS

**sensors/manufacturer/*
= man_name_select**

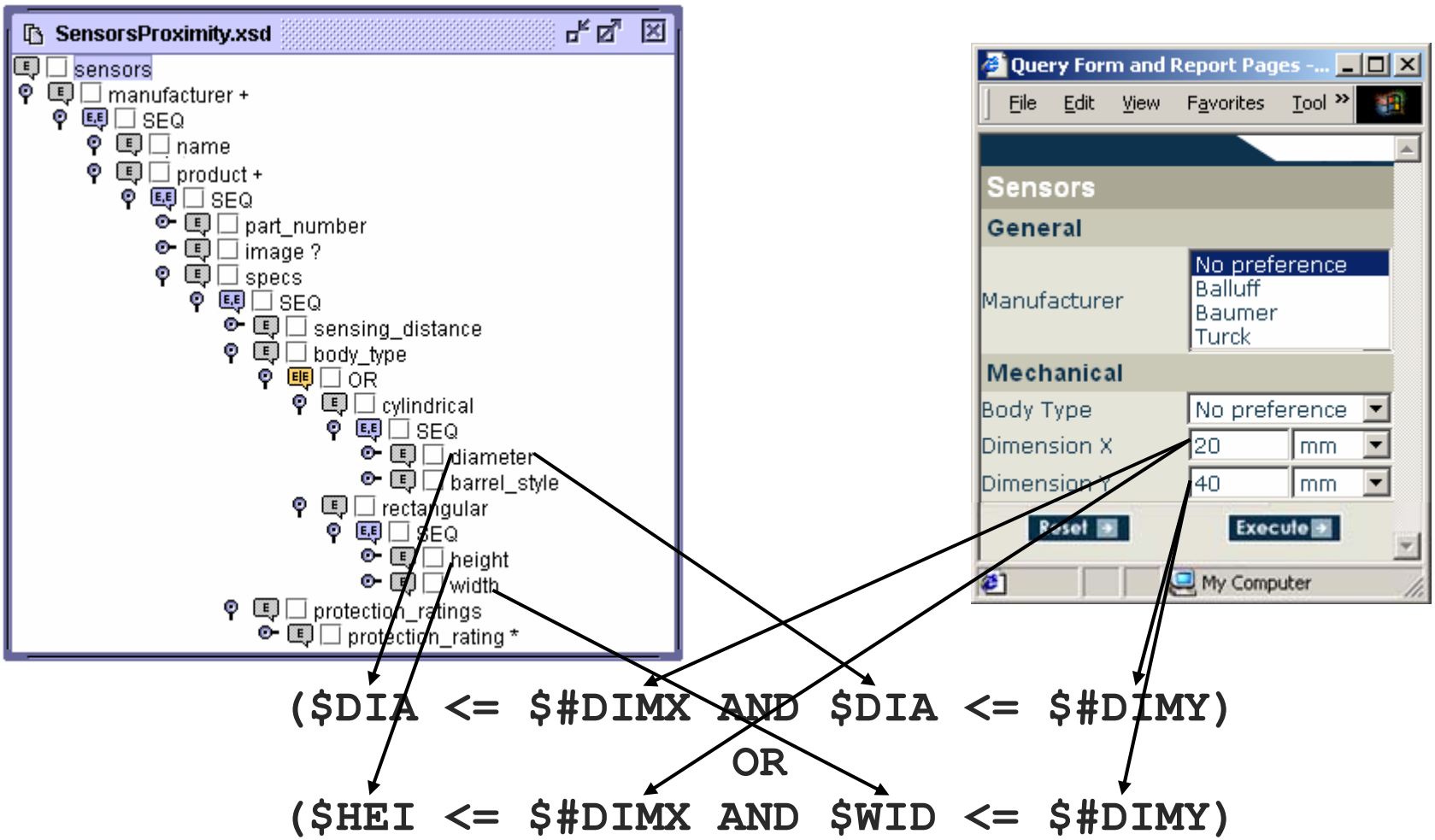


Choice of schema element e means

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

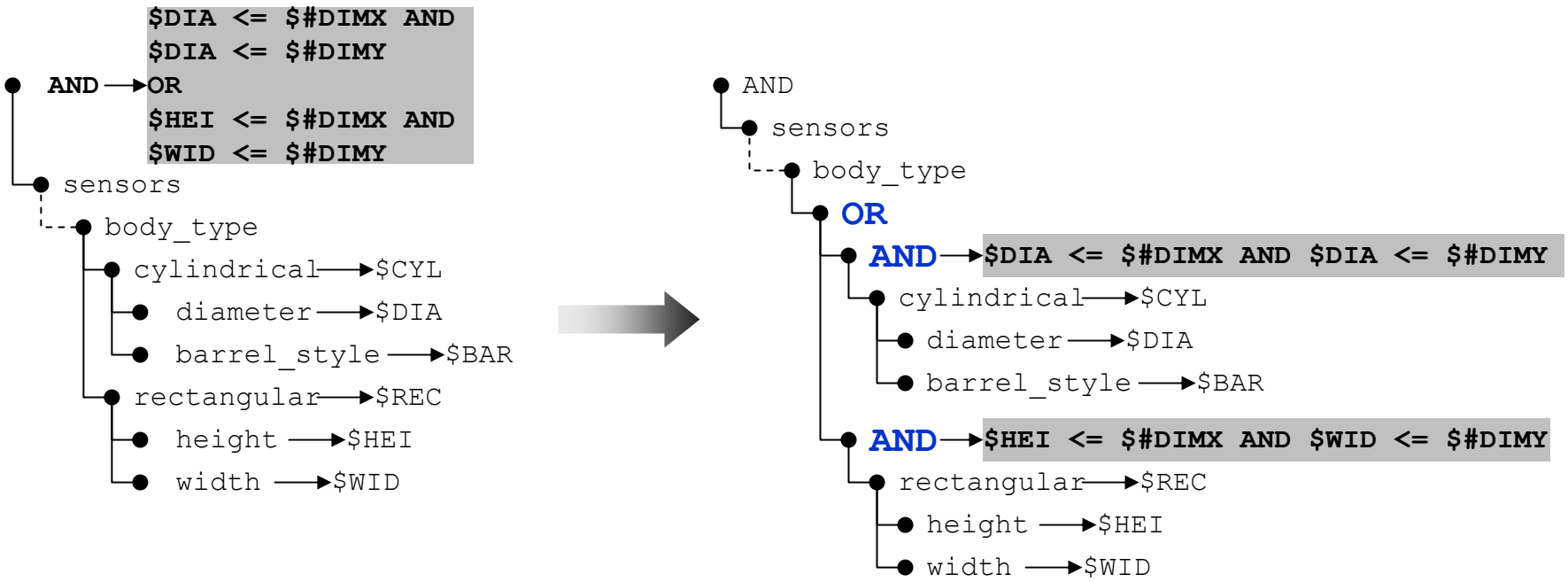
QURSED Editor

Disjunction



QURSED Editor

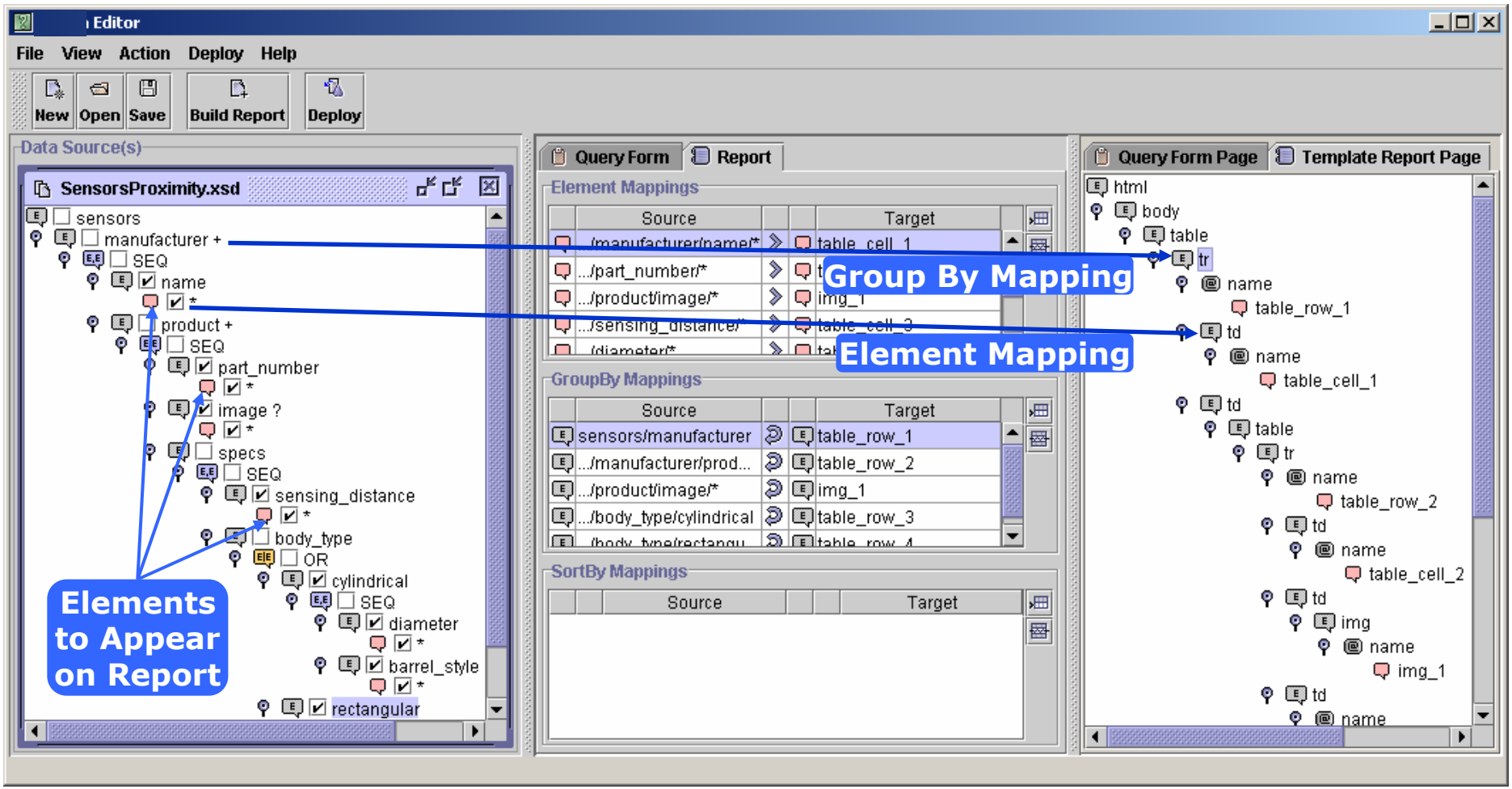
Disjunction



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

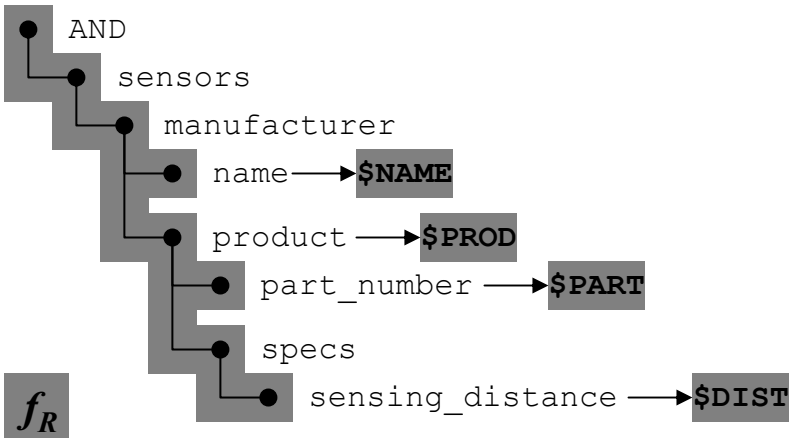
QURSED Editor

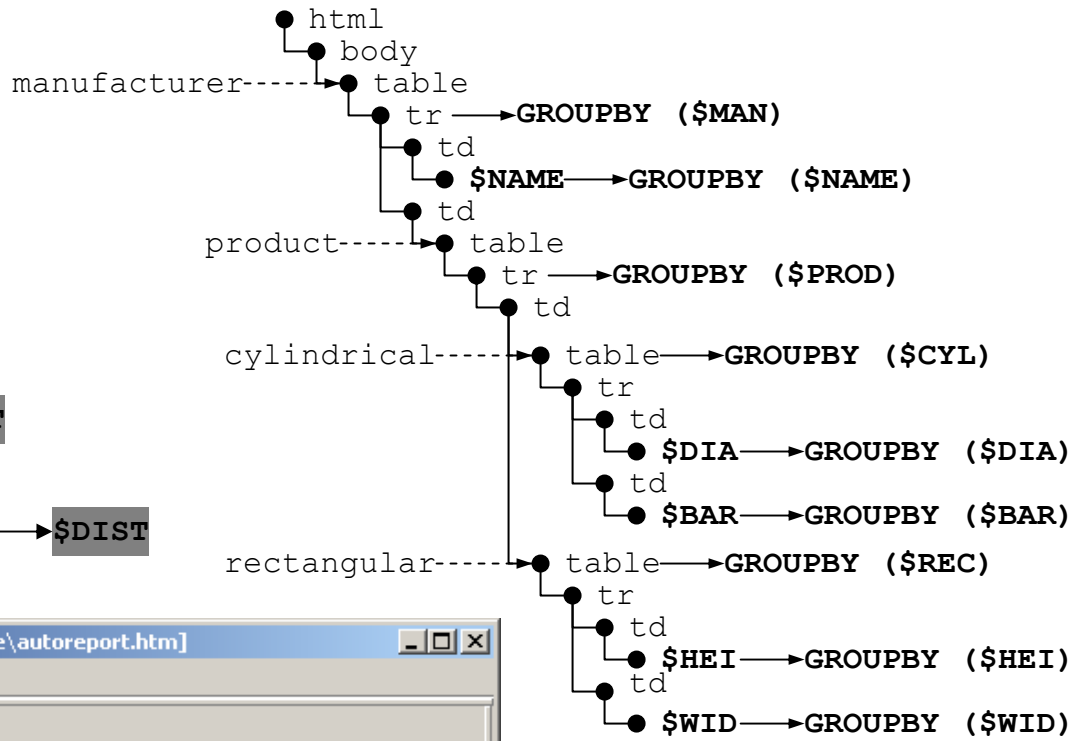
Building Reports

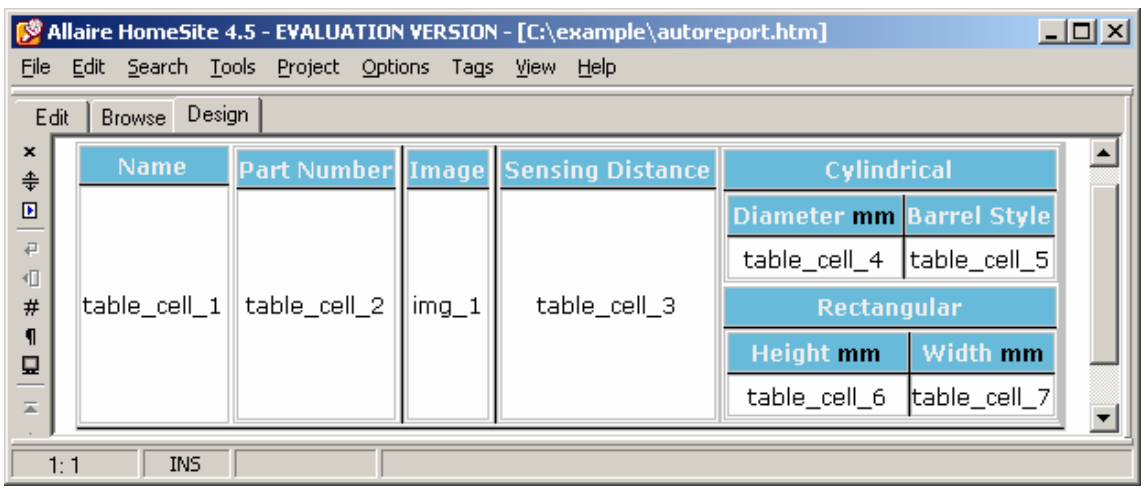


QURSED Editor

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