Model Based Mediation With Domain Maps

• Xiaosen Li (xsli2@uic.edu)
• Guanrao Chen (gchen@cs.uic.edu)
• William Sunna (wsunna1@uic.edu)

• Instructor (Prof. Isabel Cruz)
• The University of Illinois at Chicago
Outline

• Introduction
• XML-based Mediation
• Model-based Mediation
• Model-based Mediation with Domain Maps
• Application in Bioinformatics
• ISIS
• Comparison of ISIS and Model-based Mediation
Different Schemes

Federated Databases
- One-World

XML-Based Mediation
- One-/Multiple-Worlds

Model-Based Mediation
- Complex Multiple-Worlds
Our Goal

• Given different data sources:

\[ S_1 \quad S_2 \quad \ldots \quad S_n \]

• And we have different queries:
  \((Q_1,Q_2,\ldots, Q_k)\) over \((S_1,S_2,\ldots,S_k)\)

• Find answers to these questions:
  \((A_1,A_2,\ldots,A_k)\)
Introduction

Model-Based Mediation:
Integration of different data sources to retrieve information that cannot be retrieved using one source.

Domain Maps (Ontology's):
Glue Knowledge Sources
“One Simple World” example

- Given: car Dealer A, Car Dealer B
- Solution: we can use XML-Based Mediation to find the answer.
XML-Mediator (Abstract)

You can add multiple sources.
Integrated View Definition for the Car example

**XMAS**

**XML Matching And Structuring language**

CONSTRUCT  <Cars>

  <car>
  $m1
  $p

    <make>
      $ma { $ma }
    </make>

  </car> { $m1, $p }

  </cars>

WHERE <cars.car>

  $m1 : <Manu_date />
  $p : <Price/>

  </> IN

  WRAP(“Dealer_A”)

AND< Manu_dates. Manu_date >

  $m2 : <Manu_date />

  <$> IN

  WRAP(“Dealer_B”)

AND value( $m1 ) = value( $m2 )
XMAS QUERY PROCESSING

XMAS QUERY

Translator

COMPOSITION

Rewriter, Optimizer

XMAS VIEW DEFINITION

PLAN EXECUTION
XML-Based Mediation:

- XML Models
- XML Elements
- Structural Constraints:
  - DTD (Parent, Child, Sibling)
- No classes relationships (is-a, has-a)
- No logical Domain constraints
Complex Multiple-Worlds
Navigating the multiresolution data using knowledge-based mediation with domain maps
Complex Multiple-Worlds
Strategies

• Take all the huge different databases and put them into an even larger database (warehouse)

• Or develop a system to talk to different databases and correlates the results
What is the cerebellar distribution of rat proteins with more than 80% homology with human NCS-1? How about other rodents?

System that can process the query from multiple complex world databases

Protein localization
Morphology
Neurotransmission
CaBP

Federation of Brain Data
Model-based Mediation

User/Client

CM Integrated View

Integrated View Definition
IVD(S_1,S_2,..,S_k)

Mediator

CM Wrapper

XML Wrapper

CM S_1

GCM

CM S_2

GCM

CM S_k

GCM

CM Plug-ins

CM Queries & Results

S_1

S_2

S_k
Model-based Mediation

• “Lift” from syntax level to conceptual level
• Lift:
  – before: the source has element names that are NOT related
  – after: the element names are linked to a domain map
• Data provider adds links from raw data to domain maps
Model-based Mediation

- CM plug-in
  To make the mediator independent of CM formalism:
    -- Sources export all CM information in XML
    -- Use GCM so that the mediator no longer needs one module per CM formalism
Model-based Mediation

• CM to GCM
  GCM is a meta-model that any conceivable CM formalism can be expressed in.

• F-Logic as GCM
  --Convenience: root in knowledge representation and Object-Oriented database
  --Availability: FLORA, FLORID
A Question

• Different data sources contains different aspects of data. How to integrate them?

For example

- Extracellular
- Intracellular
- Cell membrane
- Calcium channel
- Ca++
Structural vs. Semantic Integration

• Source 1
  Physiological data of calcium current through calcium channels

• Source 2
  Immunolocalization of calcium channels

• Structurally they are isolated

• Conceptually and Semantically they are related
Domain Maps

- Domain Map = Ontology
  - definition of “things” that are relevant to your application
  - representation of terminological knowledge
  - explicit specification of a conceptualization
  - concept hierarchy (“is-a”)
  - further semantic relationships between concepts
    abstractions of relational schemas, (E)ER, UML classes, XML Schemas

- Formalisms:
  Semantic nets, Frame-logic, Description logic, ...
Domain Maps

- Formal definition
  - A finite set containing:
    - Description Logic (DL)
    - Logic rules
  - Facts expressed as edge-labeled digraphs with nodes representing concepts and edge labeled as roles:
    
    \[ C \rightarrow^{r} D : \text{if } c \text{ belongs to } C \text{ then there is some } d \text{ in } D \text{ such that } r(c,d) \text{ holds} \]
Domain Map

• Use in Model-Based Mediation
  --“Provide declarative means for specifying additional knowledge that is not present in the source but that can be used to navigate through and interrelate the multiple data sources.”
  --when used as part of the IVD, can infer knowledge or derive virtual relations
Knowledge based mediation
(Use of Domain Maps)

Using ontology maps to encode these semantic relationships
Domain Maps
The Whole Picture

Domain Map

Integrated View Definition
$\text{IVD}(S_1, S_2, \ldots, S_k)$

CM Integrated View

Mediator

User/Client

CM Wrapper

XML Wrapper

CM Queries & Results

GCM CM $S_1$

GCM CM $S_2$

GCM CM $S_k$

CM Plug-ins

S_1

S_2

S_k
XML-Based vs. Model-Based Mediation

CM ~ \{Descr. Logic, ER, UML, RDF/XML(-Schema), \ldots\}  
CM-QL ~ \{F-Logic, \ldots\}

Integrated-DTD := XML-QL(Src1-DTD,\ldots)

Integrated-CM := CM-QL(Src1-CM,\ldots)

Domain Maps

No Domain Constraints

Structural Constraints (DTDs),  
Parent, Child, Sibling, ... \[ A = (B^*|C), D \]

B = \ldots

CM ~ \{Descr. Logic, ER, UML, RDF/XML(-Schema), \ldots\}  
CM-QL ~ \{F-Logic, \ldots\}

Integrated-DTD := XML-QL(Src1-DTD,\ldots)

Integrated-CM := CM-QL(Src1-CM,\ldots)

Domain Maps

Logical Domain Constraints

A = (B^*|C), D  
B = \ldots

Classes, Relations,  
is-a, has-a, ...  

XML Elements

XML Models

Raw Data

Conceptual Models

Integrated-DTD := XML-QL(Src1-DTD,\ldots)

Integrated-CM := CM-QL(Src1-CM,\ldots)

Domain Maps

No Domain Constraints

Structural Constraints (DTDs),  
Parent, Child, Sibling, ... \[ A = (B^*|C), D \]

B = \ldots

CM ~ \{Descr. Logic, ER, UML, RDF/XML(-Schema), \ldots\}  
CM-QL ~ \{F-Logic, \ldots\}

Integrated-DTD := XML-QL(Src1-DTD,\ldots)

Integrated-CM := CM-QL(Src1-CM,\ldots)

Domain Maps

Logical Domain Constraints

A = (B^*|C), D  
B = \ldots

Classes, Relations,  
is-a, has-a, ...  

XML Elements

XML Models

Raw Data

Conceptual Models
Achieving Interoperability of Genome Databases Through Intelligent Web Mediators

Problem: There are hundreds or even thousands of biology databases, each with its own interface. Querying these databases are tedious, expensive and error prone.

Solution: Developing a database-independent, intelligent user interface using their existing query systems and architecture.
Abstraction Hierarchy of the Genome Database on the Web

GQL Example

```
accept into clustalx
(select clean(a.sequence)
from GlobalDB as g, AnimalDB as a
where g.organism = "Drosophila" and
g.source(country)="Kenya" and
g.journal like "USA" and
a.accession in
(select b.accession
from blast(AnimalDB, clean(g.sequence)) as b
where b.e-value >= 0.98))
```
LifeDB Web Browser

GQL Query G

Schema Query S or Query Mappings

Web Interface

Answer A

Response

Web Interface

XML Negotiator

probes

schema queries

test data

test data

Web Server

Web Interface

Response

Query G

Parameterized queries and responses

more databases

Data queries and responses

Global scheme

map info

Ontology

Query map info

Query Mappings

Global Schema

Schema Mappings

Generalizer

feedback loop

schema info

Interpreter

Query Processor

GQL query

Query plan

Parameterized queries and responses

Global scheme test data

probes

Global scheme test data

Global scheme

map info

Ontology
**ISIS Mediation Architecture**

*University of Bourgogne (France)*

- **ISIS**: *Interoperable Spatial Information System*
  - Integration of Heterogeneous Spatial or Geographic information system.
  - Multi-Agent Paradigm → Sharing spatial knowledge and Services.
  - Web Oriented Information System

- *Example of Geographic information systems (GIS’s):*
  - Road, Traffic Information on an area
  - Land use information
  - Population Distribution
  - Marketing research Demographics
**ISIS Mediation Architecture**

**MULTI-agent System Architecture**

- **Query Processing Agent**
- **Semantic router Agent**
- **Interface Agent**
- **Ontology Agent**

**Cooperation Bus**

- **CA**
- **Wrapper Agent**
- **S1**
- **S2**

**USER**

CA = Cooperation agent
**WRAPPER AGENT:**

~ processes OQL (Object Query Language) queries from Corresponding Cooperation Agent

Difference between SQL, OQL: refer to this suggested website: http://www-db.stanford.edu/~ullman/fcdb/spr99/lec15.pdf

~ Forwards the results to the Cooperation Agent

~ A wrapper Agent is an “Employee” of one Cooperation Agent.
  Responsive when triggered by the “boss”

~ Schemas are represented by AMUN (Multi-level data Model) objects
  Which Lacks Semantics

**COOPERATION AGENT:**

~ Contains knowledge of one source only (represented by Semantic Cooperation Objects)

~ Semantic Objects are created with the help of the semantic router agent

~ Process self initiated Queries or sub Queries initiated by other agents

~ Queries are written in terms of the local objects and passed to the wrapper
**ONTOLOGY AGENT:**

~ provides Mutual understanding of concepts between the various agents to help them work with each other without a need for a global schema
~ defines ontological set of terms to be used by the cooperation agents and the semantic router

**SEMANTIC ROUTER AGENT:**

~ To achieve communication between Cooperation agents, The semantic Router provides information about the location and identity of every Cooperation agent. Cooperation agents can participate in executing queries

**Query PROCESSOR AGENT:**

~ It identifies relevant information sources and creates an execution plan

**INTERFACE AGENT:**

~ Receives Queries from the user and pass them to one Cooperation agent.
~ reports back the results of the query to the user
~ only connected with one Cooperation agent
AMUN DATA MODEL

- used to represent schemas on both the wrapper level and the cooperation level.

Type hierarchy of AMUN

Geometry

Coordinate Geometry

Curve  Point  Surface  Solid

Line String  Polygon  Polyhedral surface

Line  Line ring
**ISIS vs Model-Based Mediation**

**With Domain Maps**

**ISIS:**
- Application: developed to integrate heterogeneous geographic systems in the first place
- Terminological Knowledge: Uses Ontology Agent
- Schemas: Represented by AMUN Data Model in all stages of mediation

**Model-Based:**
- Application: developed to integrate heterogeneous Biological data bases in the first place
- Terminological Knowledge: Uses domain maps
- Schemas: represented in different models in different stages (XML, CM, GCM)
QUESTIONS?

COMMENTS?
References

  http://www.sdsc.edu/~ludaesch/Paper/icde01.pdf

• [2].Model-Based Information Integration in a Neuroscience Mediator System, B. Ludäscher, A. Gupta, M. E. Martone, demonstration track, 26th Intl. Conference on Very Large Databases (VLDB), Cairo, Egypt, September 2000.
  http://www.sdsc.edu/~ludaesch/Paper/ssdbm00.html


  http://www.sdsc.edu/~ludaesch/Paper/mbm-research-11-2001.ppt

  http://www.cs.msstate.edu/~jamil/my-pub-papers/final-bibe.ps