Semantic Mapping of Relational Databases to Ontologies in a Mediation Setup Environment

Pieter Verheyden

Semantics Technology and Applications Research Laboratory
Vrije Universiteit Brussel

http://www.starlab.vub.ac.be
pverheyd@vub.ac.be

Friday, 03 December 2004

DB-dag 2004, Antwerp – Belgium
1. DOGMA ontology model
2. Case study
3. Ontological commitment language: $\Omega$-RIDL
4. Deploying ontological commitments for mediation
5. LinKFactory® and MaDBoKS
6. Ongoing and future work
7. Questions
An ontology is an “explicit specification of a shared conceptualisation of a certain domain” – T. Gruber

A DOGMA ontology is decomposed in:

- **Ontology base**
  - Holds plausible context-specific elementary fact types called *lexons*
  - Informally, a lexon is a binary fact type that lexically describes a (plausible) relationship between two concepts
  - Formally, a lexon is of the form: \(< \gamma \text{ term}_1 \text{ role co-role term}_2 >\)
  - Each (context,term)-pair refers to a unique concept

- **Commitment layer**
  - An ontological commitment formally defines a (first order) interpretation of (a subset of) the ontology base by:
    1. selecting (committable) lexons
    2. semantically constraining the use of the selected lexons by imposing (application-dependent) constraints
  - From practical experience in ontology modelling it is far less difficult to reach an agreement on conceptualisation than on constraints

- From practical experience in ontology modelling it is far less difficult to reach an agreement on conceptualisation than on constraints
Case Study

- **Health care sector**: access to correct and precise information in an efficient time frame is a necessity.

- **Hospital Information System**:
  - Several dispersed data sources:
    - Highly structured repositories (e.g., relational databases)
    - Structured documents (e.g., electronic patient records)
    - Free text (e.g., patient discharge notes)
  - Some data sources contain general information (e.g., patient’s registration information), while other contain specialised information (e.g., patient’s treatment details)
  - It is possible that there is no physical link between logically related data sources.
• SCOP – Semantic Connection of Ontologies to Patient data
  • Integrating medical relational databases through *semantic couplings* to an *existing* domain ontology
    • Keeping instance data at the database level
    • No altering of the ontology (base) in favour of the committing database
    • Instance data is queried at the ontology level and retrieved through mediation
  • Medical ontology: LinKBase®
    • Built and maintained by Language and Computing N.V. (L&C)
    • Extensive (2,000,000 concepts, 5,300,000 relationships)
  • Relational database: National Drug Code (NDC) Directory
    • Maintained by U.S. Food and Drug Administration (FDA)
    • Serves as a universal product identifier for human drugs
    • NDCs are 10-digit numbers that identify the labeller/vendor, product, and trade package size
- Exchanging ontological knowledge between both ontology frameworks
- Extract of result in simple table format:

<table>
<thead>
<tr>
<th>Context</th>
<th>Header Term</th>
<th>Role</th>
<th>Co-role</th>
<th>Tail Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEDICINE</td>
<td>DENTAL DRUG</td>
<td>IS_A</td>
<td></td>
<td>MEDICINAL PRODUCT</td>
</tr>
<tr>
<td>MEDICINE</td>
<td>MEDICINAL PRODUCT</td>
<td>HAS-PATH</td>
<td></td>
<td>ROUTE OF ADMINISTRATION</td>
</tr>
<tr>
<td>MEDICINE</td>
<td>MEDICINAL PRODUCT</td>
<td>HAS-INGREDIENT</td>
<td>IS-INGREDIENT-OF</td>
<td>INGREDIENT OF MEDICINAL SUBSTANCE</td>
</tr>
<tr>
<td>MEDICINE</td>
<td>ENTERPRISE</td>
<td>HAS-ASSOC</td>
<td></td>
<td>COUNTRY - STATE</td>
</tr>
<tr>
<td>MEDICINE</td>
<td>CANADA</td>
<td>IS_A</td>
<td></td>
<td>COUNTRY - STATE</td>
</tr>
</tbody>
</table>
National Drug Code Directory

- Drug classes
- Drug products
- Ingredients
- Packages
- Routes of administration
- Dosage forms
- Firms
• **Ontology Base Reference and Idea Language**

  • Defining an ontological commitment of a relational database to an ontology base close to its natural formulation

  • Main syntactic principles adopted from RIDL (1979)
    • RIDL\cns: constraint definition
    • RIDL\qu: conceptual query/update

• An ontological commitment consists of four closely linked parts:
  1. Commitment declaration
  2. Lexical interpretation layer
  3. Lexical association layer
  4. Semantic constraint layer

• Syntactic placeholder mechanism
define commitment in context MEDICINE with subsumption IS_A / []

lexical interpretations

map FIRMS.COUNTRY_NAME=CANADA on CANADA IS_A “COUNTRY - STATE” [] [HAS_ASSOC] ENTERPRISE

lexical associations

assoc FIRMS.COUNTRY_NAME=CHINA with “COUNTRY - STATE”

semantic constraints

each ENTERPRISE HAS_ASSOC exactly one “COUNTRY - STATE”

end
define commitment in context MEDICINE with subsumption IS_A / []

1. Declaring the context in which the commitment will be defined
2. Declaring role/co-role pair(s) to be interpreted as subsumption relation(s) in the commitment

**Issue**: almost no co-roles modelled in LinKBase®
→ Syntactic placeholder [] to denote a non-existing co-role
→ Serves as *null value* to be replaced when corresponding co-role is eventually modelled by an *authorised* ontology engineer
**Ω-RIDL: Lexical Interpretation Layer**

- Lexical mapping:
  - Path in relational database
  - Path in ontology base

- Path in relational database:
  - Attribute, e.g., `FIRMS.FIRM_NAME`
  - Attribute value, e.g., `FIRMS.COUNTRY_NAME=CANADA`

- Path in ontology base:
  - Recursively defined as an ordered sequence of lexons
  - Examples:

  "MEDICINAL PRODUCT" HAS-INGREDIENT
  "INGREDIENT OF MEDICINAL SUBSTANCE"

  ![Ontology Example](example.png)

  CANADA IS_A "COUNTRY – STATE" [I [HAS_ASSOC] ENTERPRISE

  PHYSICIAN HAS_ASSOC [XXX] PATIENT
Two kinds of lexical mappings:

1. Reference mapping
2. Relation mapping

Reference mapping: attribute or attribute value

map LISTINGS.DOSAGE_FORM on “MATERIAL ENTITY BY PRESENTATION SHAPE” [HAS_ASSOC] “MEDICINAL PRODUCT”

map FIRMS.COUNTRY_NAME=CANADA on CANADA IS_A “COUNTRY – STATE” [HAS_ASSOC] ENTERPRISE

Relation mapping: (concatenation of) foreign key(s)

map (LISTINGS.LISTING_SEQ_NO = ROUTES.LISTING_SEQ_NO, ROUTES.ROUTE_CODE = TBLROUTE.ROUTE_CODE) on “MEDICINAL PRODUCT” (HAS-PATH) “ROUTE OF ADMINISTRATION”
Ω-RIDL: Lexical Association Layer

- Lexical association:
  - Reference path in relational database
  - Term in ontology base
- Overcome (assumed) incompleteness of the ontology base
- Example:
  
  ```
  assoc FIRMS.COUNTRY_NAME=CHINA
  with “COUNTRY - STATE”
  ```
- Also to be seen as syntactic placeholder:
  if a corresponding lexon is eventually modelled, e.g.,
  ```
  < MEDICINE, CHINA, IS_A, , COUNTRY - STATE >
  ```
  then transform the association into a mapping
  ```
  map FIRMS.COUNTRY_NAME=CHINA
  on CHINA IS_A “COUNTRY - STATE” [] [HAS_ASSOC] ENTERPRISE
  ```
**Ω-RIDL: Semantic Constraint Layer**

- Semantic constraints:
  - Reflect (as good as possible) the rules intended by the UoD of the application (e.g., the integrity constraints)
  - Expressions of how the database constraints restrict the use of (i.e. *commit to*) concepts referred by the terms in the ontology base

- Example:
  each ENTERPRISE HAS_ASSOC exactly one “COUNTRY - STATE”

  constraints a lexon interpreted by:

  map FIRMS.COUNTRY_NAME=CANADA
  on CANADA IS_A “COUNTRY - STATE” [] [HAS_ASSOC] ENTERPRISE

  and reflects the attribute FIRMS.COUNTRY_NAME not allowing null values
Mediation

- Translating a conceptual query (query on ontology level) into a correct logical query (query on database level)
- Reusing the construction of ontological paths to formulate a conceptual query
- Deploying ontological commitments for the translation
- Graphically representing an ontological commitment as a tree by adopting the ORM diagram notation
Conceptual Query Example

• **Informal:**
  list all cities in Germany in which enterprises are located that are related to medicinal products having an oral route of administration

• **Formal:**
  list CITY [] [HAS_ASSOC] ENTERPRISE
  (HAS_ASSOC GERMANY and
   [] [HAS_ASSOC] “MEDICINAL PRODUCT”
   HAS-PATH “ORAL ROUTE”)

• **Graphical:**
  Sub-tree of the ontological commitment tree
Conceptual Query Example (cont.)
SELECT FIRMS.CITY
FROM FIRMS, LISTINGS, ROUTES, TBLROUTE
WHERE (FIRMS.COUNTRY_NAME = "GERMANY"
AND (FIRMS.FIRM_SEQ_NO = LISTINGS.FIRM_SEQ_NO
AND LISTINGS.LISTING_SEQ_NO = ROUTES.LISTING_SEQ_NO
AND ROUTES ROUTE_CODE = TBLROUTE.ROUTE_CODE
AND TBLROUTE.ROUTE_NAME = "ORAL")

--- left selection branch
--- middle condition branch
--- right condition branch
LinKFactory® and MaDBoKS

MaDBoKS

MetaData
Config file

XML
Mapping Structure

DB\textsubscript{x}
DB\textsubscript{y}

DBwrapper\textsubscript{x}
DBwrapper\textsubscript{y}

"Mediation layer"

mediator\textsubscript{x}
mediator\textsubscript{y}

LF server interface
Business logic
Data access layer proxy

WorkBench®

LinKFactory®

LinKBase®
LinKFactory® and MaDBoKS (cont.)

Database element

Possible matching ontology concepts
LinKFactory® and MaDBoKS (cont.)

**Blue:** concepts from the DB

**Black:** ontology concepts

**Blue:** relations between DB concepts

**Purple:** relations between DB concept and ontology concept
LinKFactory® and MaDBoKS (cont.)

Query on ontology level

Result of query
Ongoing and Future Work

- Refinement of $\Omega$-RIDL and commitment layer
- Incorporating the $\Omega$-RIDL compiler ($omegaridlc$) into DOGMA Studio ontology framework
- Possible adaptation by existing mediation technology (L&C’s MaDBoKS system)
Acknowledgments:

- IWT (SCOP project – IWT O&O #020020/L&C)
- Tom Deray, Jean-Luc Verschelde – Language and Computing N.V.
  http://www.landcglobal.com

Thank you for listening!
Questions?