Topic Maps Constraint Specification Languages

Giovani Rubert Librelotto

giovani@unifra.br

José Carlos Ramalho

jcr@di.uminho.pt

Renato Preigschadt

rpa.renato@gmail.com

Pedro Rangel Henriques

prh@di.uminho.pt
• Motivation
• Our work’s background
• What’s going on (models)?
• Constraints...
• XTche, ASTMA and OSL
• Case study
• Conclusions
Motivation

• Why do we need constraints?
  • Topic Maps are too abstract!
    ‣ Many ontologies have a rigid structure
    ‣ Many ontologies have semantic constraints
  • we do not control the process of creation
    ‣ uploads
    ‣ merges
  • manual editing of the Topic Map
• As a side effect we can have a Reporting Service
Ontology

Structure level

Catalog level
More questions

- Is a Schema a Constraint Language?

- Creating a new language ... Why?
  - We are always creating new DSL arguing that we will hide complexity
  - Aren’t we introducing complexity?

- How should we implement this new language?
Back in 2001 we had great expectations towards Topic Maps
When the number of applications grows you start to worry about the model and storage...

- In 2005 an Msc Work delivered a Relational Model for TM and a browser working over it.
- In 2006 TMDiscovery was tested:
  - Worked well with small Topic Maps
  - It was used in the classroom and in small projects
  - Failed to scale with large Topic Maps (>50,000 topics): AGROVOC, Emigration Museum DB.
• Digital Preservation
• Digital Objects: text documents, still images, relational databases.
• Model specification:
  – Descriptive Metadata (EAD)
  – Preservation Metadata (PREMIS)
  – Technical Metadata (NISO MIX, DBML, ...)
  – Wrapping metadata (METS)

http://roda.iantt.pt
RODA: the model
RODA: the model
RODA: the model
RODA: the model

First real digital object:
- 2000000 EAD nodes
- 2000000 PREMIS nodes
- 2000000 file nodes
- each preserv. event can add 2000000 nodes
RODA: the model

First real digital object:
- 2000000 EAD nodes
- 2000000 PREMIS nodes
- 2000000 file nodes
- each preserv. event can add 2000000 nodes
Returning to constraints

• What is really a TM?
  • According to the XTM DTD
    ‣ A list of ... Topics, associations, occurrences
    ‣ Everything else is constructed by reference
    ‣ It is easy to get lost in a TM!
  • That DTD is completely abstract
    ‣ There is room for “Specific Semantic Networks Schemas”
TMCL: Implementation wishes

• Do it simple
• Do it with existing technology
• Do it user friendly
  • If possible do not force the user to learn a new language
Schema constraints:
Topic of type T must have a specified number of explicit names/occurrences/subject-indicators (cardinality);
Topic of type T must have as name/occurrence/subject-indicators a value matching a particular pattern;
Topic of type T must (not) have a name/occurrence with scope S;
Topic of type T must have a name/occurrence, that is instance of topic type T, in scope S;
...

Contextual constraints:
Topic T can (only/not) be used for typing other topics;
Topic T can (only/not) be used for typing subject indicator;
Topic T can (only/not) be used for typing basenames;
Topic T can (only/not) be used for typing occurrences;
...
How to ...?

- Schematron/XCSDL processing model has been used in similar contexts
- For prototyping this processing model we would only need open "standards":
  - XML
  - XSLT
  - XML Schemas
Multi-layer XSL Processing Architecture

- Process Rules
  - XML Doc.
  - XSL Stylesheet
    - Meta stylesheet
  - XSL Proc.
  - XSL Stylesheet
    - Stylesheet that runs the process

- Instância documental
  - XML Doc.
  - XSL Proc.
  - XSL Stylesheet

- Results
  - Doc. XML
  - Pipeline

Topic Maps 2007
Pipeline

Multi-layer XSL Processing Architecture

Process Rules

- XML Doc.
- XSL Stylesheet
- Meta stylesheet

Instância documental

- XML Doc.
- XSL Stylesheet
- Stylesheet that runs the process

Some applications:
- Schematron / XCSL
- Xpath Wrapper
- Document Composer
- XDBT: XML Database Transformations

Results
XTche architecture
• XML Schema based
  • Constraint lang. = Schema lang.
  • We have some experience with XML Schemas
  • The tools we use have an interesting graphical support that we wanted to use
  • If we choose an existing language to host ours part of the work is already done

• **XTche Spec. = XML Schema + semantic stamps**
XTche Validation Process

XTCHE → XSD.XSD → ERROR → XTche.XSD → OK → XTCHE Specification Validation Processor → V
• Namespace: `xmlns:xtche=http://www.di.uminho.pt/~gepl/xtche`

xtche-schema.xsd

- A set of flags / semantic stamps
- Each flag represents a semantic validation
- The user will associate flags to concrete topics and these will map to semantic constraints that will be verified
- Remaining: Did we cover the whole TMCL?
• **Schema constraints**
  - `<xs:attribute name="topicType"/>`
  - “this element represents a Topic type”

• **Contextual constraints**
  - `<xs:attribute name="associationTypeExclusive"/>`
  - “this topic can only be used to qualify associations”

• **Existence constraints**
  - To guarantee the existence of at least a specific topic or association
A small case study

Customers, products, and orders
Restricting an Association structure:

∀a ∈ Associations: if(Type(a) = “is-making-order”)
then MemberRoles((order, product, quantity) a)
Each person must have an age field:
\[ \forall t \in \text{Topics}, \text{TopicType}(t) = \text{"person"} \Rightarrow \exists r \in \text{Occurrences} : r \subseteq t \land \text{OccType}(r) = \text{"age"} \]
Contextual constraint

Restricting the use of a specific topic: topic “wine” can only be used as a Topic Type
Semantic Stamp = xtche:associationType-Forbidden

- topic “descendant” cannot be used to type associations
- topic “marriage” cannot be used to type associations

Expected result: number of descendant and marriage associations
Astma language set

• Non-XML syntax
• Language for: structuring, adding, querying, constraining...
• An object algebra...
• It has its own model. The Topic Map must be Uploaded to this model.
forall [ * (is-making-order) 
product: * 
quantity: * 
order: * ]
forall [ * (person) 
age: * ]

Astma does not distinguish between Topics, Associations, Occurrences, ...
Ontopia Schema Language

- XML Schema based
- Allows the user to specify a schema for a specific Topic Map
- Since we are in the schema domain we are not allowed to use instance values
Comparing the 3 Languages

<table>
<thead>
<tr>
<th></th>
<th>XTche</th>
<th>ASTMA</th>
<th>OSL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Background</strong></td>
<td>XML Schema + Stamps</td>
<td>Specific Syntax</td>
<td>XML Schema</td>
</tr>
<tr>
<td><strong>Model</strong></td>
<td>Abstract Doc. Tree (XTM)</td>
<td>Specific (should be open)</td>
<td>Abstract Doc. Tree (XTM)</td>
</tr>
<tr>
<td><strong>Range of App.</strong></td>
<td>Til now... it looks promising</td>
<td>Some doubts...</td>
<td>Only structural constraints</td>
</tr>
<tr>
<td><strong>Scalability</strong></td>
<td>Does not scale</td>
<td>???</td>
<td>Does not scale</td>
</tr>
</tbody>
</table>
Conclusions

• There is work to be done
  - to optimize (in our case, in the XSL domain, it can be done a lot...)
  - to find a better model
    • Object: (id, [Property], [RelationShip])
  - or better implementations for it
Some ongoing projects (Digital Preservation):
RODA: http://roda.iantt.pt
CRiB: http://crib.dsi.uminho.pt

Questions...?

Giovani Rubert Librelotto
giovani@unifra.br

José Carlos Ramalho
jcr@di.uminho.pt