Message Validation with Semantic Reasoning Tools

Peter Denno
National Institute of Standards and Technology
Overview

- Consider the concepts of communication and its validation
- Consider approaches to validation
- Ask where the future might lead
Part 1: Communication in Engineered Systems
Q: Do these guys really know what they are talking about?

Adam

“I want to buy a foobar”

“I sell foobars. The base model is $1000”

“I need one with forged widgets”

“FOB, that is $1234”

“I’ll use Carl as my freight forwarder”

Bart

Carl

2005-04-25
Answer:

If Adam gets his foobar, and Bart and Carl get paid, yes

(or at least, who then cares?).
What matters?

- Behaviors conform to obligations between parties
- Behaviors conform to regulations and laws
- Behaviors are consistent with physical constraints
- Behaviors are cost-effective

= Best Practice
A definition:

Validation = assessing whether or not “what matters” is being achieved.
What can be validated?

- Ultimately, we want to know about “what matters”

- But between our needs and their fulfillment (Adam getting his foobar), lies…
  - expressing our needs in shared concepts
  - expressing those concepts in IT engineering detail
Syntax and Semantics

• Expressing our needs in shared concepts
  • “Message semantics” plays a role
  • Bounded only by the details of the circumstance

• Expressing those concepts in IT engineering detail
  • “Message syntax” plays a role
  • Bounded by encoded rules of “structural well-formedness”
Part 2: Approaches to Validation
Propagating Constraints

Buyer Location = Elgin, Illinois
Buyer Country = US

Seller Location = Chicago, Illinois
Seller Country = US

Domestic Sale
Incoterm = FOB

Freight Forwarder = Yellow Freight
CarrierParty = ???

Think like a prosecutor, not a programmer.

2005-04-25
OASIS Symposium - Peter Denno
Run it ‘backward’

Think like a prosecutor, not a programmer.
One way to accomplish this...

Schematron, et al.

Pro:
- Unit testing
- Elementary, proven tools

Con:
- New syntax? Throw it away.
- Repeats syntax pattern to set up context
- Domain Knowledge (DK) embedded in concerns about syntax
- Single message validation

Key

<table>
<thead>
<tr>
<th>DK</th>
<th>= domain knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>de</td>
<td>= data extraction</td>
</tr>
</tbody>
</table>
...another way

Separate Data Extraction from Domain Knowledge

Pro:
- Uncoupled from syntax
- “Unanticipated validation”
- Multi-message validation

Con:
- No unit testing
- High-tech, unproven tools
- It requires an “ontologist”
- Proof -> meaningful report
Syntax-centric approach forces the network to run in one direction

* More “procedural” means easier reporting of validation exceptions, but less circumspect, more repetitive
The Validation Tool
data extraction rules

<SYNTAX>

de
The Validation Tool

<table>
<thead>
<tr>
<th>Type</th>
<th>Information Element</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order</td>
<td></td>
</tr>
<tr>
<td>BuyersID</td>
<td>identificationSchemeVersionID</td>
</tr>
<tr>
<td></td>
<td>identificationSchemeID</td>
</tr>
<tr>
<td></td>
<td>identificationSchemeURI</td>
</tr>
<tr>
<td></td>
<td>identificationSchemeName</td>
</tr>
</tbody>
</table>

The tree depicted is the XSD, arranged as an instance document might appear, if all the elements that might occur did occur.

You can address these by Xpath. You can get the annotation provided in the UBL XSD.

2005-04-25 OASIS Symposium - Peter Denno
The Validation Tool

One Rule:

- **Top** : Documentation
- **Middle** : Condition is XPath
  (in this case, id scheme = “Incoterms”)
- **Bottom** : what is extracted
to the knowledgebase (uses XPath too)

2005-04-25  OASIS Symposium - Peter Denno
The Validation Tool

Extraction part of rule:

- Left: How extracted data is to be interpreted.
- Right: A pattern before it is edited.
Sigma KEE and Vampire used in development
http://sigma2.cim3.net:8080/sigma/KBs.jsp

Sigma knowledge engineering environment

Browsing Interface

KB Term: Buying  Show

English Word:

Buying (buying)

appearance as argument number 1

(documentation Buying "A FinancialTransaction in which an instance of CurrencyMeasure is exchanged for an instance of Physical.")

(related InternalConcept Buying Selling)

(subclass Buying FinancialTransaction)

appearance as argument number 2

Merge.txt 7906-7907

Merge.txt 7905-7905

Merge.txt 7904-7904

buying is internally related to selling

buying is a subclass of financial transaction
Part 3: The Future of Message Validation?
Future tool development effort

- Message Assembly
- Incoterms 2000
- UN Rec 16 LOCODES
- BIEs to aid data extraction
- Multi-message
- OAGIS
The future, more generally

- Multi-message validation

- Normative codes (Incoterms, LOCODES, etc) integrated with an upper ontology

- Registry content integrated with an upper ontology

- Search engine registry lookup
Conclusions

• Validation requires reasoning across a broad range of concerns.
  • “What matters,” not just XML Schema!

• Semantic reasoning tools, upper ontology, knowledge from normative codes, address that range of concerns.

• More info on the UBL tool: peter.denno@nist.gov
Backup Slides
About Context:

- **Registry Viewpoint:**
  “A context names a category of usage of a term.”

- **Ontology Viewpoint:**
  “’Context’ is the word we use to refer to the constraints on interpretation before we know them.”
About Distinctions:

- **Registry viewpoint:**
  
  “Partition vocabularies where distinctions seem likely.”

- **Ontology viewpoint:**
  
  “Show me the distinction, and I’ll encode it.”
More definitions:

- **System** = a collection of parts that work jointly toward some goal.

- **Systems integration** = enabling the communication that allows the parts to work together

- **Semantic integration** = that part of system integration that concerns whether or not messages, correctly received and disassembled, in fact serve to direct the recipient to perform the desired behavior.