3. XML Foundations; Introduction to Modeling

DE + IA (INFO 243) - 30 January 2008

Bob Glushko

Plan for Today's Lecture

XML Foundations for Document Engineering

Models and modeling

- The classical modeling approach
  - Modeling for analysis; physical models
  - Modeling for design; conceptual models
- Adapting the classical approach in Document Engineering
- Methodologies for modeling

Neither the methods of Document Engineering nor the models it produces have anything inherently to do with XML or any other syntax.

But HTML has limited use for business applications because it has no tags for marking up information to give it business meaning.

XML -- which is just 11 years old -- has become the preferred format for representing physical models of documents and business processes.

XML was designed to give the intuitive idea of a document model a more physical, formal foundation.

XML is a metalanguage for markup, and markup languages -- "document types" -- can be created for very specific document or process models.


XML schemas define the rules that govern the arrangement and values of a document’s content.

An XML schema communicates the model of a document type to people or applications that need to create or receive document instances, so an XML document without a schema is little more than a bag of tags whose meanings are undefined.

There is often a gap between the conceptual model of a document -- all the "business rules" associated with it -- and what can be described in an XML schema.

The decision about where to transform documents is a business one.
A Model of Counting or Inventory

Not A Model of Wonton Soup
A Model of Wonton Soup

**Wontons**
- ½ lb. ground chicken,
  - veal or lean pork
- 2 green onions, finely chopped
- 2 tsp. grated fresh ginger root
- 1 tsp. cornstarch (corn flour)
- 1 tsp. lite soy sauce
- 6 drops hot pepper sauce
- 18 wonton wrappers

**Soup**
- 4 cups chicken broth
- 2 tsp. lite soy sauce
- 1 tsp. sesame oil
- ½ tsp. hot chili paste
- Salt and freshly ground black pepper
- 1 cup shredded fresh spinach or watercress
- 1 green onion, thinly sliced

In bowl, combine ground chicken, green onions, ginger root, cornstarch, soy sauce and hot pepper sauce; mix until well blended. Lay wonton wrappers on work surface. Place spoonful of ground chicken mixture in center of each one.

Moisten edges with water. Fold each wrapper in half to form triangle; press edges to seal. Arrange in steamer basket. Cover and steam over boiling water for 3 minutes.

Meanwhile, in large saucepan, combine broth, soy sauce, sesame oil, and hot chili paste. Bring to boil. Season to taste with salt and pepper.

Just before serving, stir in spinach until wilted. Place 3 wontons in each bowl. Ladle soup over top. Garnish with green onion rings.

---

What is a Model?

Models are *simplified descriptions* of a subject that abstract from its complexity to *emphasize some* features or characteristics while *intentionally de-emphasizing others*

Models enable us to describe and communicate systems regardless of the specific domain or discipline that they represent

A model can represent a human activity, a natural system, or a designed system

We can model structures – objects, their characteristics, their static relationships with each other like hierarchy, and reference

We can model functions, processes, behaviors – dynamic activities that create and affect structures
Recipes as Everyday Models

A recipe describes both objects and structures (ingredients) and the processes (instructions) for creating a food dish.

Important characteristics and uses of the recipe model are:

- You can FOLLOW the recipe to create the dish.
- You can COMMUNICATE the recipe to someone else who can then create the same dish.
- You can use the recipe as a GUIDE FOR EXPERIMENTATION with the objects or the processes in the recipe to create alternative dishes.

"Static" or "Structure" Recipe Model: The Ingredients
"Dynamic" or "Process" Recipe Model: The Instructions

The Classical Modeling Approach
Physical Modeling for Analysis [1]

The primary purpose of modeling is to better understand some existing system or environment and its entities ("the things that exist in it that contain or embody relevant information") and to describe this understanding so it can be communicated

Models of things as they currently exist are *Physical* or *As-Is* models

This modeling activity is usually called "systems analysis" or simply "analysis"

A basic task of modeling for analysis is capturing the languages and practices of the people who work with the instances and artifacts in the "real world"
Physical Modeling for Analysis [2]

Any system (especially business ones) has groups of stakeholders who do not fully understand the "big picture" – an analysis model can be used to prevent or repair misunderstandings.

A physical model synthesizes different views or observations into a more complete or generic perspective that accurately accounts for all of them.
Conceptual Modeling for Design or Re-Design [1]

The next purpose of modeling is to assist in the design or re-design of a system or set of artifacts.

This modeling activity is usually called "systems design" or simply "design".

Models of things as they could be are Conceptual or To-Be models.

Design abstracts away or generalizes from the technology and implementation details in the physical model to create a conceptual model.

Conceptual Modeling for Design or Re-Design [2]

A conceptual model that is implementation-independent is often easier to talk about than one that is encoded in a specific technology context.

When technologies change, the implementation model may change but the conceptual model won't.

When implementations in different technologies are based on the same conceptual models, they can more readily understood because of their common conceptual components.
Using Conceptual Models

Objects can be manipulated as conceptual components without impacting the real world that they describe, or in ways that are impossible in the real world.

This encourages the re-use of common components via standardization, patterns and libraries.

It facilitates the rationalization of components and the removal of redundancies and inefficiencies.
The Modeling Gaps

There is an essential difference or gap between the real world being modeled and any models of it, or else the models would serve no purpose.

Likewise, there is always a gap between a physical model and a conceptual model, because an analysis model is often most useful when it isn't tied to specific or feasible implementations or technologies.

But this means we can sometimes see what the current world looks like and what we would like it to be without being able to see how to get from one to the other.

Put a different way, this means that models can be designed that are impossible to implement.

Implementing a Conceptual Model as a Physical One

A model is purely theoretical until it is encoded in a technology that lets it operate in the real world again.

This is often a two-stage process: encoding conceptual models as physical ones, and then applying transformations to create instances with desired properties (e.g., implementing / formatting / rendering in some concrete medium / syntax / technology).

When instances implemented in different technologies are generated or re-generated from models, they can more readily interoperate because of their common conceptual components.
Applying the New Conceptual Model of Wonton Soup

Iteration is Inevitable
Adapting the Classical Modeling Approach to Document Engineering [1]

In Document Engineering context we can analyze a domain and analyze its entities and their processes or behaviors as we would in any other modeling activity.

But what we generally care more about is the information or intangible content about the entities in the domain.

Or put another way, we only model those entities and their properties that convey information that is relevant.

So we are more likely to engage in "pseudo-real-world" modeling than "real-world-modeling".

Real-World Modeling of a Library System
Pseudo-Real-World Modeling of a Library System

So We Model the Information, not the Things

<Book>
  <Title>Document Engineering</Title>
  <Author>Glushko</Author>
  <Author>McGrath</Author>
  <Publisher>MIT Press</Publisher>
</Book>
Document Instances are External Views

...<Book>
  <Title>Moby Dick</Title>
  <Author>Herman Melville</Author>
  <Publisher>Airmont</Publisher>
</Book>

<Book>
  <Title>Document Engineering</Title>
  <Author>Robert J. Glushko</Author>
  <Author>Tim McGrath</Author>
  <Publisher>MIT Press</Publisher>
</Book>

Physical Model Encoded as XML Schema

...<xs:element name="Book">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="Title" type="xs:string"/>
      <xs:element name="Author" type="xs:string"
                  maxOccurs="2"/>
      <xs:element name="ISBN" type="xs:string"/>
      <xs:element name="Publisher" type="xs:string"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
</xs:schema>
Conceptual Model Encoded as XML Schema

...<xs:element name="Book">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="Title"/>
      <xs:element ref="Authors"/>
      <xs:element ref="ISBN"/>
      <xs:element ref="Publisher"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
</xs:element>

Methodologies – Disciplines for Modeling

[1]

When we create a model we follow – implicitly or explicitly – some steps or techniques for analysis, design, and implementation.

This is called the modeling methodology.

Methodologies can be formal, prescriptive, step-by-step, documented and auditable or they can be the opposite: informal, ad hoc, "seat of the pants" with no trace other than the model artifact itself.
Sequential, Iterative, and Artifact-Centered Methodologies

A methodology's process describes the work to be done and the order in which it is to be done (the modeling workflow)

Many methodologies prescribe a Sequential process -- the "waterfall" model

Other methodologies are more iterative or recursive -- like the "spiral" model of progressive refinement or "agile" modeling

Other methodologies are looser about the modeling activities but emphasize the results that must be obtained at each step or phase

Methodologies – Disciplines for Modeling

A methodology can contain or define:

- Meta-models
- Processes / Activities / Steps
- Notations
- Tools

and how these are applied or fit together to produce:

- Artifacts
Readings for 4 February

"Beyond CPFR" Paul Demery, Internet Retailer (February 2006)

"Paperless Trading: Benefits to APEC", Australian Department of Foreign Affairs and Trade, (2001)

"Adoption of UBL in Denmark – business cases and experiences" M. Brun, J. Brown and R. Lohde, XTech (2005)