XML Schemas

Lecture 10, 07/10/02

What are XML Schemas

- Evolving standard for defining document structure and content model.
- More powerful and versatile (though complex at times) than DTDs.

Why Schemas

- XML Schemas are extensible to future additions
- XML Schemas are richer and more useful than DTDs
- XML Schemas are written in XML
- XML Schemas support data types
- XML Schemas support namespaces

Purpose of XML Schemas (and DTDs)

- Specify:
  - the structure of instance documents
  - "this element contains these elements, which contains these other elements, etc"
  - the datatype of each element/attribute
  - "this element shall hold an integer with the range 0 to 12,000" (DTDs don't do too well with specifying datatypes like this)

Highlights of XML Schemas

XML Schemas are a tremendous advancement over DTDs:
- Enhanced datatypes
  - 44+ versus 10
  - Can create your own datatypes
  - Example: "This is a new type based on the string type and elements of this type must follow this pattern: ddd-dddd, where 'd' represents a digit."
- Written in the same syntax as instance documents
- Less syntax to remember
- Object-oriented
  - Can extend or restrict a type (derive new type definitions on the basis of old ones)
  - Can express sets, i.e., can define the child elements to occur in any order
  - Can specify element content as being unique (keys on content) and uniqueness within a region
  - Can define multiple elements with the same name but different content
  - Can define elements with nil content
- Can define substitutable elements - e.g., the "Book" element is substitutable for the "Publication" element.

Acknowledgements

A great portion of this presentation has been borrowed from Roger Costello’s excellent presentation on XML schema that is available online at http://www.xfront.com/xml-schema.html
What do DTDs provide us with?

What do Schemas provide us with?

All XML Schemas have "schema" as the root element.
XMLSchema Namespace

http://www.w3.org/2001/XMLSchema

complexType

element

sequence

complexType

element

string

boolean

integer

This is a directive to any instance documents which conform to this schema.

An element used by the instance document which were declared in this schema must be namespace-qualified.
Referencing a schema in an XML instance document

```xml
<?xml version="1.0"?>
<BookStore
    xmlns="http://www.books.org"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.books.org BookStore.xsd">
    <Book>
        <Title>My Life and Times</Title>
        <Author>Paul McCartney</Author>
        <Date>July, 1998</Date>
        <Publisher>McMillin Publishing</Publisher>
    </Book>
</BookStore>
```

1. Referencing a schema in an XML instance document

A schema defines a new vocabulary. Instance documents use that new vocabulary.

XMLSchema-instance Namespace

http://www.w3.org/2001/XMLSchema-instance

BookStore.xsd

- defines elements in namespace http://www.books.org

Note multiple levels of checking

BookStore.xsd

- defines elements in namespace http://www.books.org

Validate that the xml document conforms to the rules described in BookStore.xsd

Validate that BookStore.xsd is a valid schema document, i.e., it conforms to the rules described in the schema-for-schemas

Validate that BookStore.xsd is a valid schema document, i.e., it conforms to the rules described in the schema-for-schemas