Database support for XML

Tore Risch
Information Technology
Uppsala University

What is XML?

• Originally HTML subset of SGML
• HTML text markup language
• XML larger subset than HTML
• HTML has predefined markup tags, e.g. `<a...>` ... `</a>`
• XML allows programmer to define and use user defined tags
• As in HTML annotation of document sections also allowed:
  `<a html="#UI">User interface</a>`

DTD

• Grammar of XML documents, e.g. allowed tags, can be described (constrained) by DTD-documents
• Can be seen as simple schema
• Original purpose still document markup
• XML documents don’t require DTDs

XML for data exchange

• There is a lot of need to exchange data between system, e.g. by e-mail or between programs
• DTDs allow to define standard formats (schema) for data to exchange
• E.g. record structures, lists, etc.
• DTDs defined e.g. for various business exchanges
• Still clumsy format
Example XML data

```xml
<?xml version "1.0" standalone="no">
<!doctype projects system "proj.dtd">
<projects>
  <project>
    <name>Toys</name>
    <number>1</number>
    <worker>
      <ssn>123456</ssn>
    </worker>
  </project>
  ...
</projects>
```

Example DTD

```xml
<!doctype projects [  
  <!element projects (project+)>  
  <!element project (name, number, workers)  
  <!element workers (worker+)>  
  <!element worker (ssn)>  
]>  

Schema for XML documents. Unlike RDBs sequences not sets. ‘any’ might be specified for subtree
Semistructured data!
```

Problems with XML for data exchange

- DTDs are voluntary and need not be followed
- Only datatype is `string`
- XML still very common and standard schemas for various application areas have been defined

XML-Schema

- XML-Schema is extension of DTDs
- XML syntax for schema too
- Very rich set of built-in data types
- User defined data types
- Inheritance
- Can be seen as object-oriented schema
- Complex type system
- Oracle XML DB can translate XML-Schema -> OR schema for subsequent querying
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <xsd:element name="Person">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="SSN" type="xsd:integer"/>
        <xsd:element name="WorksAt" type="Department" maxOccurs="1"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>

  <xsd:element name="Department">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="Dname" type="xsd:string"/>
        <xsd:element name="Workers" type="Person" maxOccurs="unbounded"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
</xsd:schema>

XML query languages

- XPath, based on regular path expression selecting XML substructures of tree structure:
  doc(user.it.uu.se/~torer/doc)/Persons/Worker/SSN

- XQuery also allows joins and constructing new document views, FLWR expressions:

  for $x in doc(user.it.uu.se/~torer/doc)/Person where $x/Name eq Toys return <ssn> $x/Worker/SSN </ssn>