Design and Implementation of XML Database Applications

Shuangbao Wang, Ph.D.
Table of Contents

- XML and Databases
- Make Decision: Relational, XML or NXD
- KIHd System – an XML Database Application
- Data Security
XML

- **XML**: Extensible Markup Language
  - **Extensible**, unlike HTML
    - Users can add new tags, and *separately* specify how the tag should be handled for display
  - **Goal**
    - To replace HTML as the language for publishing documents on the Web *
- **XML** has become the basis for all new generation data interchange formats
  - **Examples:**
    - Banking: funds transfer
    - EDI: Order processing (especially inter-company orders)
    - Scientific data
Storage of XML Data

- XML data can be stored in
  - Non-relational data stores
    - Flat files
      - Natural for storing XML
      - no concurrency, no recovery, …
    - XML database
      - Database built specifically for storing XML data, supporting DOM model and declarative querying
  - Relational databases
    - Data must be translated into relational form
    - Advantage: mature database systems
    - Disadvantages: overhead of translating data and queries
XML Database

- **XML-Enabled Databases**: Databases with extensions for transferring data between XML documents and themselves.
  - Oracle, ms SQL Server, DB2, Access etc.
- **Native XML Databases**: Databases that store XML in "native" form, generally as some variant of the DOM mapped to an underlying data store.
  - Middleware: Software you call from your application to transfer data between XML documents and databases.
    - Microsoft ADO, mysql, IBM, Oracle etc.
- **IDEs and Editors**: Software designed to help you write XML applications or edit XML documents.
  - XML spy, Delphi etc.
Why use XML Database

- Rule-based markup
- Not merely to show up on a web, it can be understood in any XML-ready environment
  - Acrobat, IE, Flash MX, Dreamweaver MX, ms Office
  - PDA, cell phone, meta-data
- Extensible language, can be customized to suit the need of any institutions or businesses.
- Web Service – XML over HTTP
  - an application delivered as a service that can be integrated with other Web Services using Internet standards.
  - it's an URL-addressable resource that programmatically returns information to clients who want to use it.
  - The major communication protocol used is **SOAP**, the Simple Object Access Protocol, which in most cases is XML over HTTP.
Modern Web

Structure
- HTML
- XHTML
- XML

Presentation
- CSS1
- CSS2

Behavior
- ECMAScript
- DOM
Relational Database

- Relational or XML?
  - Relational Database
    - Oracle
    - ms SQL server
    - DB2
    - Access
    - File Maker
  - XML
    - XML enabled Database
    - Native XML Database (NXD)
Database Application Development

E-R  TABLE

Entity Relationship Access
**T/TAC Online**

- Virginia's Training and Technical Assistance Centers (T/TAC)
  - to improve educational opportunities and contribute to the success of children and youth with disabilities (birth - 22 years).
  - provide quality training and technical assistance in response to local, regional, and state needs.
  - increase the capacity of schools, school personnel, service providers, and families to meet the needs of children and youth.
  - [http://ttaconline.org](http://ttaconline.org)
E-R Diagram – Training @ T/TAC Online
Relational Database

**StaffProfile**
- username
- password
- firstName
- lastName
- phone
- email
- title
- employer
- region1
- region2
city
state
zip
category

**Courses**
- username
- sTitle
- sTitle1
- summary
- minutes
- category
- category2
- subcat1
- subcat2
- status
- sDate

**Webshops**
- username
- sTitle
- summary
- minutes
- category
- category2
- subcat1
- subcat2
- status
- sDate

**WebshopPages**
- username
- sTitle
- pid
template
gTitle
status
toxc1
ga1
pCaption1
pSource1
pAlt2

**Images1**
- Username
- sTitle1
- pTitle1
- id
- pid
template
filename
filesize
contenttype
filedata

**Relationship**

Dr. Shuangbao Wang

Kellar Institute
George Mason University
# Data in Relational Database

<table>
<thead>
<tr>
<th>username</th>
<th>sTitle</th>
<th>summary</th>
<th>minutes</th>
<th>category</th>
<th>category2</th>
<th>subcat1</th>
<th>subcat2</th>
<th>status</th>
<th>sDate</th>
</tr>
</thead>
<tbody>
<tr>
<td>ajohnson</td>
<td>Using Intellitalk @ajohnson1114</td>
<td>This workshop will take place...</td>
<td>25</td>
<td>2</td>
<td>6</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>11/20/2004</td>
</tr>
<tr>
<td>ajohnson</td>
<td>Using Word Predict ajohnson1127</td>
<td>This workshop will be held...</td>
<td>25</td>
<td>2</td>
<td>6</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>11/27/2004</td>
</tr>
<tr>
<td>ajohnson</td>
<td>How to Use Scanners ajohnson1212</td>
<td>This workshop will take place...</td>
<td>25</td>
<td>2</td>
<td>6</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>12/1/2004</td>
</tr>
<tr>
<td>ajohnson</td>
<td>How to Make a Tall ajohnson8112</td>
<td>Many students learn...</td>
<td>25</td>
<td>2</td>
<td>6</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>8/11/2004</td>
</tr>
<tr>
<td>ajohnson</td>
<td>Introduction to Int ajohnson8122</td>
<td>This workshop will be held...</td>
<td>20</td>
<td>2</td>
<td>6</td>
<td>15</td>
<td>0</td>
<td>-1</td>
<td>8/12/2004</td>
</tr>
<tr>
<td>ajohnson</td>
<td>Using Intellitalk@ajohnson8132</td>
<td>This workshop will take place...</td>
<td>25</td>
<td>2</td>
<td>6</td>
<td>15</td>
<td>0</td>
<td>-1</td>
<td>8/13/2004</td>
</tr>
<tr>
<td>ajohnson</td>
<td>Using Intellitalk@ajohnson8132</td>
<td>This workshop will take place...</td>
<td>25</td>
<td>2</td>
<td>6</td>
<td>15</td>
<td>0</td>
<td>-1</td>
<td>8/13/2004</td>
</tr>
<tr>
<td>ajohnson</td>
<td>Tour of Intellitalk@ajohnson8142</td>
<td>This workshop will be held...</td>
<td>20</td>
<td>2</td>
<td>6</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>8/15/2004</td>
</tr>
<tr>
<td>ajohnson</td>
<td>Using Intellitalk3 ajohnson8920</td>
<td>This workshop will be held...</td>
<td>25</td>
<td>2</td>
<td>6</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>8/13/2004</td>
</tr>
<tr>
<td>allison</td>
<td>Transition Planning allison6220041</td>
<td>This workshop offer...</td>
<td>25</td>
<td>2</td>
<td>6</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>8/13/2004</td>
</tr>
<tr>
<td>AmieF</td>
<td>&lt;h3&gt;History and Planning amief7132004</td>
<td>Overview...</td>
<td>10</td>
<td>6</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>1</td>
<td>6/24/2004</td>
</tr>
<tr>
<td>AmieF</td>
<td>&lt;h3&gt;Roles &amp; Responsibilities amief7272004</td>
<td>This workshop focus...</td>
<td>25</td>
<td>6</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>10/1/2004</td>
</tr>
<tr>
<td>AmieF</td>
<td>&lt;h3&gt;Knowledge Cl amief7920048</td>
<td>The knowledge...</td>
<td>10</td>
<td>4</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>7/27/2004</td>
</tr>
<tr>
<td>AmieF</td>
<td>&lt;h3&gt;History and Planning amief7920048</td>
<td>Overview...</td>
<td>10</td>
<td>6</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>1</td>
<td>10/1/2004</td>
</tr>
<tr>
<td>AmieF</td>
<td>&lt;h3&gt;Activities amief7920049</td>
<td>Work with the supc...</td>
<td>10</td>
<td>6</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>10/1/2004</td>
</tr>
<tr>
<td>AmieF</td>
<td>&lt;h3&gt;Related Resources amief7920049</td>
<td>The following are rec...</td>
<td>5</td>
<td>6</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>10/1/2004</td>
</tr>
<tr>
<td>AmieF</td>
<td>&lt;h3&gt;References - amief8320040</td>
<td>This is a list for all rec...</td>
<td>5</td>
<td>6</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>10/1/2004</td>
</tr>
<tr>
<td>AmieF</td>
<td>&lt;h3&gt;Families and Children amief8920042</td>
<td>Families are wono...</td>
<td>25</td>
<td>6</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>10/1/2004</td>
</tr>
<tr>
<td>bcarter6</td>
<td>Effective Bullying P bcarter672220</td>
<td>Dr. Bonnie Bell Cart...</td>
<td>28</td>
<td>3</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>7/22/2004</td>
</tr>
<tr>
<td>coquina77</td>
<td>Lever Switch Made coquina77711</td>
<td>This workshop demc...</td>
<td>28</td>
<td>2</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>8/29/2004</td>
</tr>
<tr>
<td>coquina77</td>
<td>Tilt Switch Made fro coquina77715</td>
<td>This workshop Expl...</td>
<td>28</td>
<td>2</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>8/25/2004</td>
</tr>
<tr>
<td>coquina77</td>
<td>Pillow Switch Made coquina77716</td>
<td>This workshop demc...</td>
<td>25</td>
<td>2</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>8/29/2004</td>
</tr>
<tr>
<td>coquina77</td>
<td>On/Off Switch Made coquina77719</td>
<td>This workshop demc...</td>
<td>25</td>
<td>2</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>8/26/2004</td>
</tr>
<tr>
<td>elander0</td>
<td>Introduction to Overelay ul elander05122</td>
<td>Overlay Maker is a...</td>
<td>25</td>
<td>2</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>1</td>
<td>5/21/2004</td>
</tr>
<tr>
<td>elander0</td>
<td>Introduction to Overlay Overlay Maker is a...</td>
<td>25</td>
<td>2</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>1</td>
<td>5/13/2004</td>
<td></td>
</tr>
<tr>
<td>elander0</td>
<td>SW Programs for Relay relay072220</td>
<td>Estela Landeros, T...</td>
<td>25</td>
<td>2</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>7/22/2004</td>
</tr>
</tbody>
</table>

Dr. Shuangbao Wang

Kellar Institute

George Mason University
Kellar Instructional Handheld data (KIHd) System

- Using Personal Digital Assistants to Collect Discrete Trial Data
- Increase in Severe Disabilities, including Autism
- Federal Emphasis on Access to Curriculum
- Teacher and School Accountability
- Student Performance
- Maturing Technologies
Datatype of KIHd System
## XML Datatype Schema

### datatype
- **name**: frequency
- **yesno**: answer
- **counter**: counts
- **name**: duration

#### loop
- **sn**: number
- **start**: time1
- **stop**: time2

#### name: accuracy

#### loop
- **sn**: number
- **yes**: answer1
- **no**: answer0

#### name: fluency

#### loop
- **sn1**: number1
- **start**: time1
- **stop**: time2

#### loop2
- **sn2**: number2
- **yes**: answer1
- **no**: answer2
XML Database Schema

datatype

frequency
- yesno
- counts

duration
- loop

accuracy
- loop
  - sn
  - start
  - stop
  - sn
  - yes
  - no

fluency
- loop
  - sn
  - start
  - stop
  - sn2
  - yes
  - no
XML file of the datatype
XML Document Schema

- Database schemas constrain what information can be stored, and the data types of stored values
- XML documents are not required to have an associated schema
- However, schemas are very important for **XML data exchange**
  - Otherwise, a site cannot automatically interpret data received from another site
- Two mechanisms for specifying XML schema
  - **Document Type Definition (DTD)**
    - Widely used
  - **XML Schema**
The type of an XML document can be specified using a DTD. DTD constraints structure of XML data:

- What elements can occur
- What attributes can/must an element have
- What subelements can/must occur inside each element, and how many times.

DTD does not constrain data types:

- All values represented as strings in XML

DTD syntax:

- `<!ELEMENT element (subelements-specification) >`
- `<!ATTLIST element (attributes) >`
XML Schema

- XML Schema is a more sophisticated schema language which addresses the drawbacks of DTDs. Supports
  - Typing of values
    - E.g. integer, string, etc
    - Also, constraints on min/max values
  - User defined types
  - Is itself specified in XML syntax, unlike DTDs
    - More standard representation, but verbose
  - Is integrated with namespaces
  - Many more features
    - List types, uniqueness and foreign key constraints, inheritance ..
- BUT: significantly more complicated than DTDs
Querying and Transforming XML Data

- Translation of information from one XML schema to another
- Querying on XML data
- Above two are closely related, and handled by the same tools
- Standard XML querying/translation languages

  - **XPath**
    - Simple language consisting of path expressions
  - **XSLT**
    - Simple language designed for translation from XML to XML and XML to HTML
  - **XQuery**
    - An XML query language with a rich set of features
A **stylesheet** stores formatting options for a document, usually separately from document

- E.g. HTML style sheet may specify font colors and sizes for headings, etc.

- The **XML Stylesheet Language (XSL)** was originally designed for generating HTML from XML

- XSLT is a general-purpose transformation language

  - Can translate XML to XML, and XML to HTML

- XSLT transformations are expressed using rules called **templates**

  - Templates combine selection using XPath with construction of results
XQuery

- XQuery is a general purpose query language for XML data
- Currently being standardized by the World Wide Web Consortium (W3C)
  
  The textbook description is based on a March 2001 draft of the standard. The final version may differ, but major features likely to stay unchanged.
- Alpha version of XQuery engine available free from Microsoft
- XQuery is derived from the Quilt query language, which itself borrows from SQL, XQL and XML-QL
- XQuery uses a
  
  \( \text{for } \ldots \text{ let } \ldots \text{ where } \ldots \text{ result } \ldots \) syntax
  
  \( \text{for } \Leftrightarrow \text{ SQL from} \)
  
  \( \text{where } \Leftrightarrow \text{ SQL where} \)
  
  \( \text{result } \Leftrightarrow \text{ SQL select} \)
  
  \( \text{let } \) allows temporary variables, and has no equivalent in SQL
XML in KIHd System

- In Kellar Instructional Handheld Data system (KIHd), data and presentation are separated and data is shared between different computers, PDAs, or even cell phones.
- A portable “recordset” is used and is to be interpreted by different devices through XSL.
- This also adds security to the system database so that client can only “view” portion of the data they requested.
Teacher enters a login screen to identify the person who will be collecting the data.

Teacher identifies “student” and desired instruction, including “domain”, “skill area” and specific “skill objective” or so called “task” to working on.

Teacher identifies behavioral prompts to be utilized from a checkbox list.

Teacher selects the data type to be used in the instructional approach.

Teacher selects Phase and start the Session to begin collecting data.

Teacher collects data on individual student performance. PDA confirms selections and actual data is collected and displayed on the PDA screen.
Graph representation of student learning curve

- The last 10 sessions of student behavior data are graphed on the PDA
- Help teacher better understanding student behaviors and his/her learning curves.
Augmented Reality and Monitoring

- On head mounted devices or 3D displays, the image of the background classroom is overlaid by the target scene (child).
- A personalized view of learning environment is displayed.
- Parents can only be able to see their child in the classroom.
- They monitor their child behaviors so that they can assist teaching the child at home.
Nearly 10 million people were victimized by identity theft last year. The lost reached 5 billion.

In early March, the nation’s largest data miner ChoicePoint with 19 billion data files include driver’s license, ssn, credit history, birth certificate, real estate deed, and even thumbprint and DNA was broken into and some 145,000 people’s data was extracted.

“Our system of protecting people’s identity is virtually nonexistent in this country.” - Senator Charles Schumer

His staff was able to download personal information on the likes of Dick Cheney and Brad Pitt from a ChoicePoint rival, Westlaw.

The damage caused by intrusions far beyond money and privacy, it also becomes the real threat to the national security.
DB Application System Structure
XML Web Service

XML Web Service

User Interface

Web Client Application

Web Form

Dataset Authors1

DataGrid

Business Object

XML Web Service

Public Interface
GetAuthors
Update Authors

Private Members
Dataset authors1.xsd
SqlDataAdapter
SqlConnection
SqlCommand

Data

OLEDB

pubs

Windows Client Application

Windows Form

Dataset Authors1

DataGrid

XML Web service call via HTTP

XML Web service call via HTTP
SOAP

- stands for **Simple Object Access Protocol**
- a communication protocol
- for communication **between applications**
- a format for **sending messages**
- designed to communicate **via Internet**
- **platform independent**
- **language independent**
- based on **XML**
- simple and extensible
- allows you to **get around firewalls**
- will be developed as a **W3C standard**
A portable “recordset” is used and is to be interpreted by different devices through XSL.

- When the first time a query command is issued, a portable “recordset” is generated and xml web service verifies the data
- If the time elapses for retrieving the data exceeds the threshold or
- If the data requested is more than the number of trails (10)
- XML Web Service will refuse to deliver the recordset to clients
- Otherwise, data are sent to clients via SOAP
Search

- Alphabet.doc (26,112 bytes)
- Alphabet.txt (34 bytes)
- Search alphabet.doc
- Search XML Document
<?xml version="1.0" encoding="UTF-8" ?>

<datatype>
  <name>frequency</name>
  <yesno>answer</yesno>
  <counter>counts</counter>
  <name>duration</name>
  <loop>
    <sn>number</sn>
    <start>time1</start>
    <stop>time2</stop>
  </loop>
  <name>accuracy</name>
  <loop>
    <sn>number</sn>
  </loop>
  <name>fluency</name>
  <loop>
    <loop2>
      <sn1>number1</sn1>
      <start>time1</start>
      <stop>time2</stop>
      <loop2>
        <sn2>number2</sn2>
        <yes>answer1</yes>
        <no>answer2</no>
      </loop2>
    </loop2>
  </loop>
</datatype>
Conclusion

- Use XML enabled databases in development of database applications
- XML is system independent
  - Windows, Unix, Linux
- Cross-platform access
  - Ubiquitous – Computer, PDA, Cell Phone
- Exchange data
  - EDI
- Searching the data is easy and efficient
  - Search engines can simply parse the description-bearing tags rather than muddling in the data. Tags provide the search engines with the intelligence they lack.
Questions and Comments

Shuangbao Wang, Ph.D.

paul.wang@computer.org

http://virtual01.gmu.edu