

XML and Web Services

Susan B. Davidson
572 Levine North, 898-3490,
susan@cis.upenn.edu

<http://www.cis.upenn.edu/~susan/cis650/homepage.html>



Fall 2004

CIS 650



1

Pre-requisites

- An introductory database class (e.g. CIS550).
- Some knowledge of XML and XQuery is useful, although a rapid introduction will be given this week and next.
- Lack of inhibition when it comes to asking questions, expressing your viewpoint, and presenting papers.



Fall 2004

CIS 650



2

The revolution

- **Jim Gray:** "Databases have become the vehicles to deliver an integrated application development environment, to be data-rich nodes of the Internet, to do data discovery and to be self-managing."
- **Serge Abiteboul:** "XML and Web services are revolutionizing the automatic management of distributed information, somewhat in the same way that HTML, Web browsers and search engines modified human access to world wide information."



Syllabus (tentative)

- Sept. 9:** Introduction to the topics, overview of the course.
- Sept. 14-16:** Overview of XML, XQuery and Web Services
- Sept. 21-23:** Relational database storage techniques for XML
- Sept. 28-30:** XQuery to SQL translation
- Oct. 5-7:** Native XML storage and query optimization
- Oct. 12-14:** Active XML
- Oct. 19-21:** E-Services and security
- Oct. 28 - Nov. 4:** Web service composition
- Nov. 9-11:** Updates to XML
- Nov. 16-18:** XML change management
- Nov. 23:** XML views and updates
- Nov. 30- Dec. 9:** Project presentations.



How the class will run....

- You will each will pick a small number (2-3) papers to present sometime during the semester
 - I will work with you on the presentation!
- Each student who is not presenting will prepare a 1-page summary of the paper with questions
 - I expect students to be interactive and to ask questions!
- You will also do a project, which can be either research oriented or an implementation

- This class is meant to explore current database research, so if there is a topic/paper not on the syllabus that you'd like to see included, talk to me.



The Web today

- Protocol: HTTP
- Documents: HTML
- Millions of independent web sites and billions of documents
- Browsing and full-text indexing
- Publication of databases using forms



Two recent developments

- XML: syntax for describing data
 - Model: labeled ordered trees
 - Moves from a document world to a data world
 - Standard for data exchange
- Web services
 - Ability to activate a method on a remote Web server
 - Ubiquitous XML distributed computing infrastructure



The vision: *Looking for a truc*

1. Query a directory (yellowpages): who is a Truc provider?
2. Negotiate with the candidate providers
 - Nature of service
 - Quality/cost/etc.
3. Use the information
 - Get the information
 - Introduce the service in your processing
4. Eventually compose services
5. Eventually publish services



Web services

- Next step in the evolution of the WWW.
- Allow active objects to be placed on Web sites providing distributed services to potential clients.
- Uses:
 - E-commerce
 - Management of distributed information
- Propose web solutions that can be easily deployed across machines, operating systems and application languages



Components of infrastructure

- **SOAP: Simple Object Access Protocol.**
 - XML-based lightweight protocol for information exchange in a distributed environment
 - Allows XML types of arguments and service results to be specified
 - Can be used with HTTP
- **WSDL: Web service Definition Language**
 - Language for describing web service interfaces
 - Analogous to Corba's IDL for the web
 - XML format for describing network services based on operations and messages
 - The abstract description is then bound to a concrete network protocol and message format to define functionality



More components

- UDDI: Universal Discovery Description and Integration
 - Specification for distributed Web-based information registries of Web services
 - Publicly accessible set of implementations
- WSDL: Web Service Flow Language
 - XML language for the description of service compositions
 - Specify usage patterns (business processes) or interaction patterns (partner interactions)



The starting point: XML

- Blends schema and data into one format
 - Unlike relational model, where we need schema first
 - ... But too little schema can be a drawback, too!
- Allows us to get data from all sorts of things
 - Allows us to touch data we don't own!
 - This was actually a huge change in the DB community
- Interesting relationships with DB techniques
 - Useful to do relational-style operations
 - Leverages ideas from object-oriented, semistructured data

