

# XML, the Web, and Database Functionality?



Ling Liu  
Associate Professor  
College of Computing  
Georgia Institute of Technology

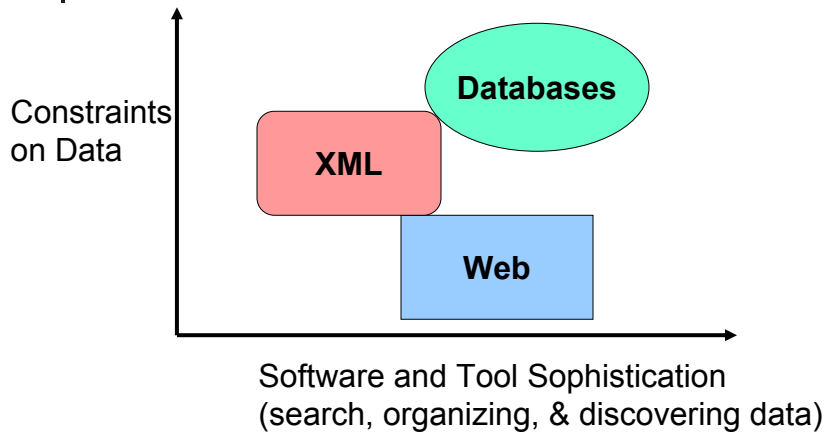


## In Real Life

- What grows the fastest in the Cyberspace?
  - Data
- How much data is there?
  - LOC: a few petabytes, 3.5million sound recordings: a few petabytes, 500,000 movies in DVD: a few petabytes  
[Michael Lesk97, JimGray98]
- Which technology is most popular for data transport?
  - XML
- What is the most successful technology today for storing and managing terabytes or petabytes of data?
  - DB Technology



## XML, Web, Database



If we consider constraints as one way to express semantics about data, XML can be seen as adding more constraints to the data.



## Database People Too Far?

- No ...
- Positive: DB people know the Truth
  - Imposing constraints allow stronger software
- Current DBMSs manage HTML/XML data
  - Insufficient constraints lead to non-applicability of the best of DBMS functionality
  - ***It's not Database People's fault***

## Database People Too Far? (cont.)

- **Challenge:** database guys should know the Truth
  - Web data don't fit into rigid database schemas, cannot be managed by DBAs, ...
  - Web data cannot reside in one database
  - Serious applications need interoperability, consistency, security, ...
- **Current web tools have DB/transactions**
  - WebSphere and WebLogic
  - challenges: easy to use, lower price, more automation, strong support for interoperation ...

## What is the big deal about XML

- XML: A Wire Protocol for Data Exchange
- Simple, Self-describing
- human readable
- easy to generate & easy to maintain
- portable and carries well
- easy to represent a hierarchy of concepts
- bridge the divide of various wire interoperation protocols
- **For users**
  - XML powered Web will be faster, friendlier, a better place to do business
- **For tomorrow's Web designers**
  - XML is a more robust infrastructure for searching, organizing, and discovering data
    - the second generation Web



## Is XML good for everything?

- Bad for data storage
  - can relational DBs be used for storing XML data?
- Bad for sending massive data over the wire fast
  - poor performance: 10x common
  - need efficient encoding/compression mechanisms
- Bad for handling updates
  - poor consistency
  - view updates are hard
    - classical issues on view updates (forward/backward)
- Not sufficient for interoperability
  - Are ontologies the complimentary solutions?



## Research Challenges

- DB People
  - Mechanisms for Efficient XML storage and Fast XML transportation
  - XML view updates
- IR and DB People
  - Scalable, Friendlier, higher quality XML search tools
    - XML Enabled Wrappers, XML-enabled Crawlers, Indexers, and Ranking Algorithms ...
- KM and DB People
  - More automated mechanisms for encoding semantics & ontologies



More ...

