Sharing Design Information

H. V. Jagadish
University of Michigan
Talk Outline

- Motivation – The role of information in the design process.
- XML
- Schema
- Current Work
Design Information Flows

Designers
Design Information Flows

Designers

Requirements, feedback, ...

Customers
Design Information Flows

Designers

Suppliers, Co-vendors,…

Customers
Design Information Flows

Designers

Suppliers

Co-designers

Customers
Design Information Flows

Customers

Suppliers

Designers

Co-designers

ασδε{ ΙΛ - καθηΓ

Customers

Co-designers

Suppliers

Designers

Design Information Flows
Need for Standards

• Communication Protocol -- TCP/IP
• Character Encoding – ASCI, Unicode
• Syntax – XML
• Schema -- Industry (vertical) segment specification??
Talk Outline

• Motivation
• XML
• Schema
• Current Work
• XML is the de facto standard for data interchange today.
• XML is block-structured -- each start tag has a matching end tag, with start/end pairs nesting properly.
• XML document corresponds to a tree, with tag nesting corresponding to edges.
A Structured XML Fragment

<doc>
  <title> Pooh’s List </title>
  <list>
    <item> eeyore </item>
    <item> tigger </item>
  </list>
</doc>
Annotation Benefits

Can tag data, at any (sub-tree) granularity, with annotations with regard to:

- data pedigree -- how and when obtained
- access control -- who can read/write/edit
- interpretative comments -- e.g. “Results of simulation includes 1999 model of part XYZ. All other parts use most current version.”
XML with Annotations

Diagram:
- Document
  - Commentary
    - Title
    - List
      - Creator
      - Pedigree
        - Item
        - Item
Query Processing with Annotations

- Convert annotations, where appropriate, into conjunctive predicates in query.
- Build indices for such annotation predicates, when possible
  - E.g., Authorization index
- Perform query optimization, and predicate re-ordering with standard cost-based techniques.
Talk Outline

- Motivation
- XML
- Schema
- Current Work
Need Schema

- Standards can help, but cannot solve the entire problem.
- Can avoid gratuitous differences
  - E.g. “horse-power” vs. “HP” vs. “horsepower”
- But need room to evolve
  - Want low barrier to introduction of new kinds of data, but the new material may not be standardized…
Schema Differentiation

- Schema differences will arise due to:
  - Evolution
  - Mergers and acquisitions
  - New cooperation, changing alignments of players in a dynamic marketplace.

- Still need to be able to communicate – query and respond.
Discover Schema

• Self-describing structures carry schema information with them.
• Can be used to process queries.
• Query may need run-time modification as schema is discovered.
• Schema representation structures, and their use for querying, are both active areas of research.
Talk Outline

- Motivation
- XML
- Schema
- Current Work
  - Overall Research Program
  - In Aid of Mechanical Design
How to Manage XML Data

• Quick route to commercialization
  – Build a small translation layer on top of an existing relational DBMS.
  – Adopted by all current major DB vendors.
  – But has significant performance and modeling disadvantages.

• Desired eventual objective
  – Native XML database.
Difficulty with Relational Map

Make up a single relation

Books
Title, Auth, Price, Pages, ISBN, ...

Socks
Size, Color, Price

TVs ...

Object_Id,
Attribute_Name,
Value_of_Attr
Research Issues

• Physical Organization
  – Assignment of data to disk
  – Compression of queriable fields

• Logical Organization
  – Formal data model
  – New Index Structures

• Query Optimization
Previous Practice

Invocation with specific parameter values

Advisor → TDES

Results in a proprietary format
Our Idea

Tool X

Advisor

Tool Y

XML

TDES
End Result

Tool X

Advisor

Tool Y

XML DB

TDES
For each (attempted) invocation of TDES by Advisor
If run with same parameter values in database
Then return results of old run.
Else really invoke TDES
Lesson Learned

• Good software design for interoperability across tools/organizations/processes also leads to better reuse even with a single designer/tool.
Conclusion

• Information technology is revolutionizing the way we do business, and strongly impacting the way we (can) do design.
• Continuing with the old ways is a sure way to fail in the new world.
• At the ARC, we have begun to address these issues; but there is much left to be done in harnessing new comp. technologies.
The Corporate Approach to Data Sharing

- Carefully developed standards
- Painfully negotiated schema
- Vigorous enforcement
- Efficient but brittle
- Change is difficult -- live with old choices
Complex Corporate Structure

- Mergers and acquisitions
  - information systems merger is a big barrier to exploiting synergies
- Some companies are run like holding conglomerates -- need divisional autonomy yet need global reporting
- Franchising corporations need to have an arm’s length relationship with franchisees.
Hierarchical Database Solution

• Structure database hierarchically
• Each sub-tree can be autonomous
  – may have different schema
• Parent node can be queried based on “greatest common information” of children.
• Wrappers at parent can permit more expressive querying.