XML
eXtensible Markup Language

- User-defined tags: Metadata
- Tags give structure – hierarchy
  - Must be nested
- Tags have attributes

  <atag anattribute=“avalue”>

Specification

- XSL – eXtensible Style Language
  - How display information
- DTD = Document Type Declaration
  - User specifies own tags and attributes
  - User-defined grammar for syntax
- XML schema – similar to but more general than DTD

- XML “salvation” for searching dynamic objects and non-text objects

- Tags say what object is

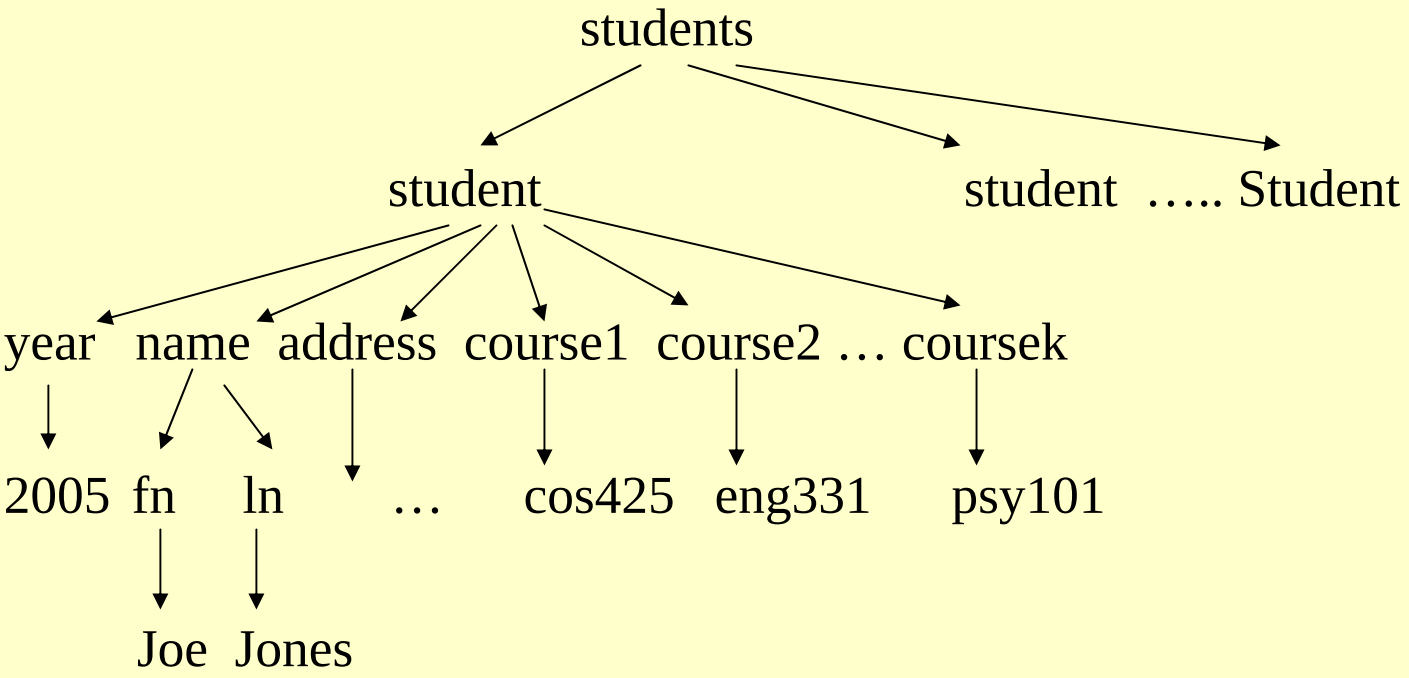
=> return to “library model” – cataloging?

Semistructured Data Model

- XML gives structure, but not fully or rigidly specified
- Use labeled graph model
  - For XML get trees
    - Node: atomic or compound object
    - Leaves: values

Example

```
<students>
    <student>
        <year>2005</year>
        <name><fn>Joe </fn><ln>Jones</ln></name>
        <address>…</address>
        <course type="deptal">cos 425</course>
        <course type="elective">eng 331</course>
        etc.
    </student>
    <student> ……</student>
        ….
</students>
```


students
student
student ….. Student
year
name
address
course1
course2 … coursek
2005
fn
ln
…
cos425
eng331
psy101
Joe
Jones

Querying XML

- Storing data in XML; want to query
- Could map to relational model, but then must restructure data
- Several querying languages

XQUERY

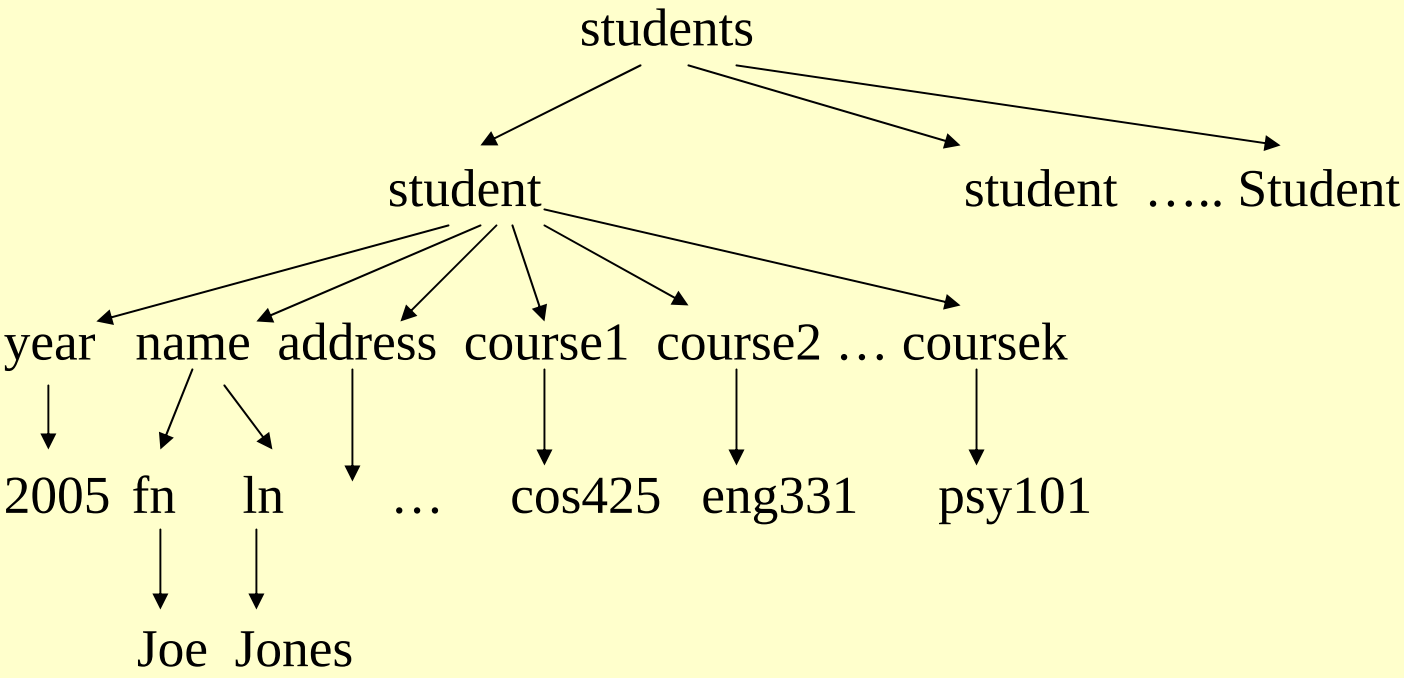
- Take a brief look at XQUERY
  - FOR *path expression* – anal. to SQL "FROM"
  - LET *path expression*
  - WHERE *condition* – anal. to SQL "WHERE"
  - RETURN – returns XML doc.

    Anal to SQL "SELECT"
- XQUERY returns XML fragment

Path expression

Doc. Identifier	//	/
e.g. URL	indicates element	indicates immed.
root of tree	nested anywhere-	child of path so
	jump down tree	far
	at this point in path	

Returns set of matching elements


students
student
student ….. Student
year
name
address
course1
course2 … coursek
2005
fn
ln
…
cos425
eng331
psy101
Joe
Jones

EXAMPLES

FOR $x IN doc_id//name/ln
RETURN <Result>$x</Result>

Gives: <Result><ln>Smith</ln></Result>
<Result><ln>Jones</ln></Result>
etc.

FOR $b IN doc_id/students/student
WHERE $b/year=‘2005’
RETURN <Result>$b/name/fn, $b/name/ln </Result>

Gives: <Result><fn>Joe</fn><ln><Jones></ln></Result>
<Result> …. etc.

Beyond XML: Semantic Web

From W3C Semantic Web Activity Statement:

http://www.w3.org/2001/sw/Activity

"The goal of the Semantic Web initiative is as broad as that of the Web: to create a universal medium for the exchange of data. It is envisaged to smoothly interconnect personal information management, enterprise application integration, and the global sharing of commercial, scientific and cultural data."

Overview

- Initiative of W3C: World Wide Web Consortium - of academic, government and industry
  - begun 1994 by Tim Berners-Lee
- provides common frameworks for data specification allowing sophisticated functionality
  - Allowing automated understanding and use of information
- Two major frameworks:
  - Resource Description Framework (RDF),
  - Ontology Language (OWL)
- Open specifications, open source
  - Allow independently written tools interoperate

RDF

- Graph model to represent *resources* and relationships between them
  - Documents and other resources
- Formal semantics
- XML syntax
- URIs for naming –Uniform Resources Identifiers
  - Generalization of URLs
- Update released Feb 2004

RDF representation

- Represents “Web resources”
  - Documents on Web
  - Generalizes to “objects” identifiable but not directly retrievable, e.g. shopping facility
- Represents metadata for resources
  - Information about information
    - Title, author, copyright of document
  - Price, shipping date of an item for sale

RDF Graph Model

- Nodes: resources and property values
- Edges: labeled with property identifiers (i.e. attribute names)

Example from http://www.w3.org/TR/rdf-primer/

"there is a Person identified by http://www.w3.org/People/EM/contact#me, whose name is Eric Miller, whose email address is em@w3.org, and whose title is Dr. "

Example: Graph representation

Nodes: 1. http://www.w3.org/People/EM/contact#me

2. http://www.w3.org/2000/10/swap/pim/contact#Person
3. Eric Miller
4. mailto:em@w3.org
5. Dr.

Edges:

1 -> 2 labeled http://www.w3.org/1999/02/22-rdf-sytax-ns#type

1 -> 3 labeled http://www.w3.org/2000/10/swap/pim/contact#fullName

1 -> 4 labeled http://www.w3.org/2000/10/swap/pim/contact#mailbox

1 -> 5 labeled http://www.w3.org/2000/10/swap/pim/contact#personTitle

Example: syntax

```
"<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#">
   <contact:Person
   rdf:about="http://www.w3.org/People/EM/contact#me">
      <contact:fullName>Eric Miller</contact:fullName>
      <contact:mailbox rdf:resource="mailto:em@w3.org"/>
      <contact:personalTitle>Dr.</contact:personalTitle>
  </contact:Person>
</rdf:RDF> "
```

OWL

- Advanced support for

  software agents

    Programs that “understand” and can plan and act

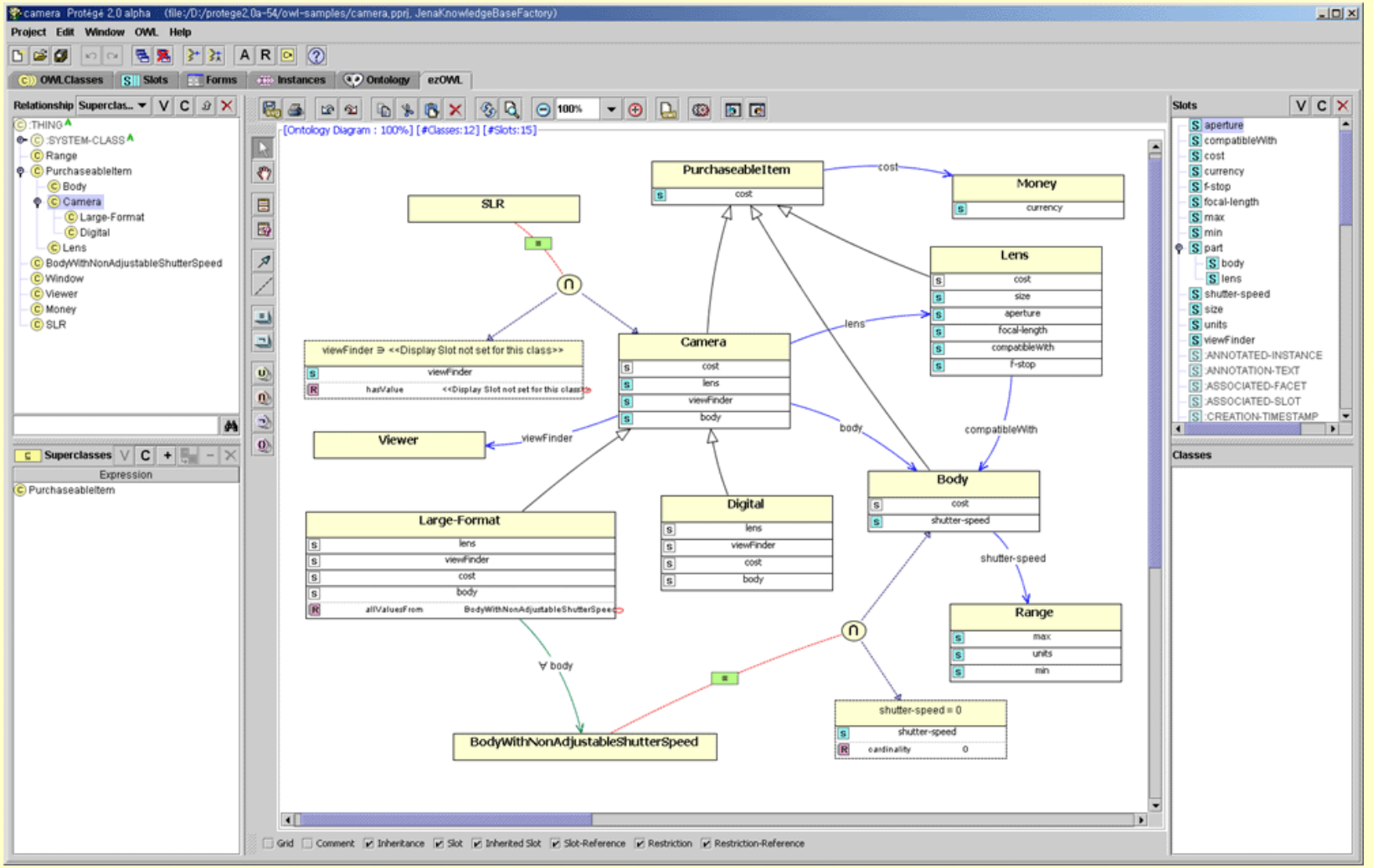
  knowledge management

    Finding and exploiting complex interactions of information across sources
- Builds on RDF
- Represents *ontologies*
- Released Feb 2004

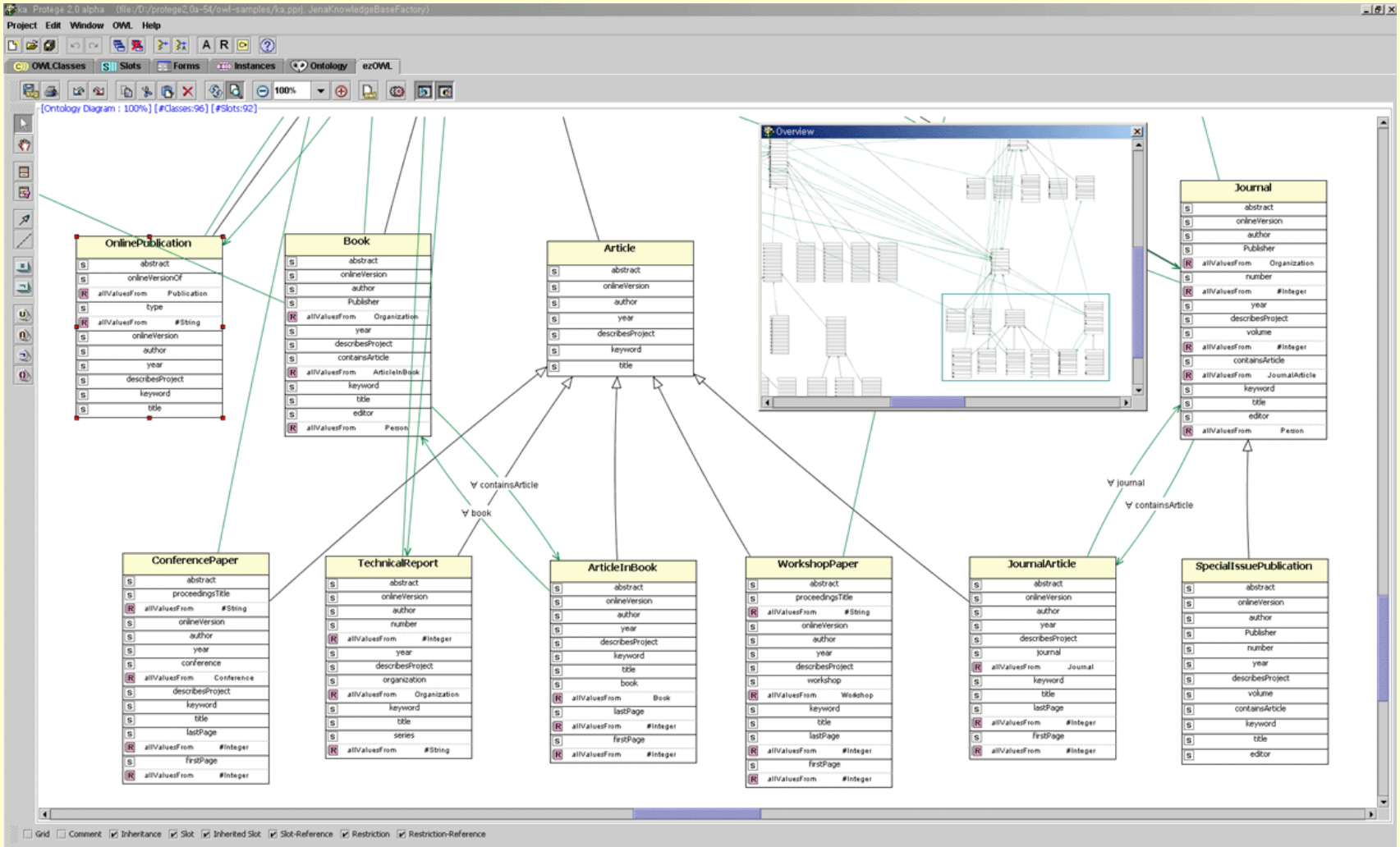
OWL expressiveness

- Ontology: “representation of terms and interrelationships”
  - very general
  - not just trees
- Has formal semantics
- Can represent relationships between classes

Example ontology



Example ontology



References

- Previous two pictures from Electronics and Telecommunications Research Institute of Korea *ezOWL* project, a Semantic Web Ontology Editor:
  http://iweb.etri.re.kr/ezowl/screenshot.html
- W3C semantic web:
  http://www.w3.org/2001/sw/
  http://www.w3.org/2001/sw/Activity
- W3C information on RDF:
  http://www.w3.org/TR/rdf-primer/
- W3C information on OWL:
  http://www.w3.org/2004/OWL/
  http://www.w3.org/TR/owl-features/
- Background on W3C:
  http://www.w3.org/Consortium/#background