Digital Libraries

The Semantic Web: Making sense of it all

Week 6

Min-Yen KAN

Motivation for semantic web

"The primary goal is to make the web more like a library and less like a heap of messy books on the floor."

-Tim Bray, Textwise consultant

- The Web can reach its full potential only if it becomes a place where data can be shared and processed by

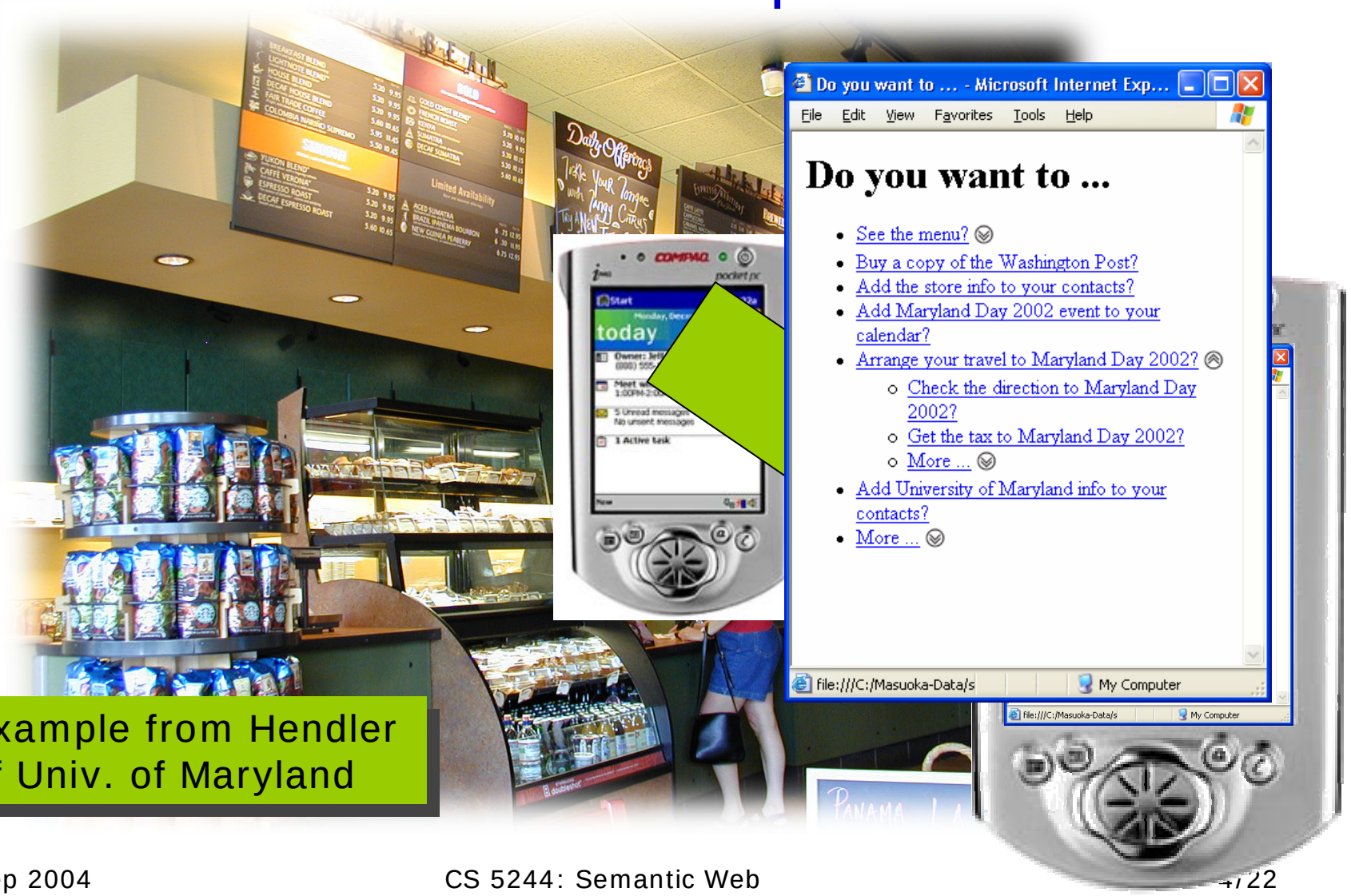
  automated tools as well as by people.

  - Semantic Web Activity Statement

So, what is it anyways?

- An extension of current the web that allows:
  - Exchange of data
    - By software agents
    - Will allow agents to reason
  - Needs to be able to seamlessly exchange data

Other examples: Services off the desktop



Example from Hendler of Univ. of Maryland

Or perhaps on different desktops…

This is great, why didn't we think of it sooner?!?

- As a community, we have been trying:
  - Doug Lenat spun off a company to try to capture commonsense knowledge in a huge knowledge representation project.
    - 1,000,000 assertions captured.
    - Cyc knows that trees are usually outdoors, that once people die they stop buying things, and that glasses of liquid should be carried right-side-up.

What's different about the Semantic Web?

CYC and the Semantic Web

- One company
- Centralized
- First-order logic
- Complex KR language
- Authoritative data

- Many companies
- Distributed
- *Left up to agent*
- Simple KR language
- Noisy data

Will this work? No one's sure.

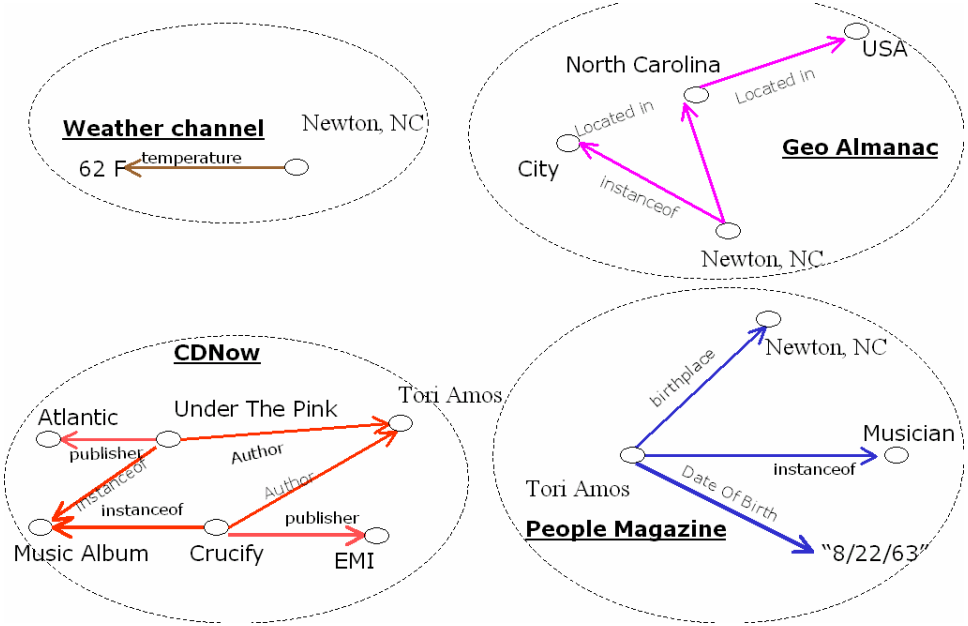
Proponents say its just a matter of time. Naysayers say we are revisiting the fundamental KR problems. BTW, CYCorp has put their ontology into the SW developmental efforts.

Implementing SW

- For the semantic web to function, computers must have access to:
  - structured collections of information
  - and sets of inference rules that they can use to conduct automated reasoning.
- Adding logic to the Web — the means to use rules to make inferences
- Uses XML and RDF as a framework

Semantic Web problem

- Islands of XML from disparate web services
- Example : Tori Amos

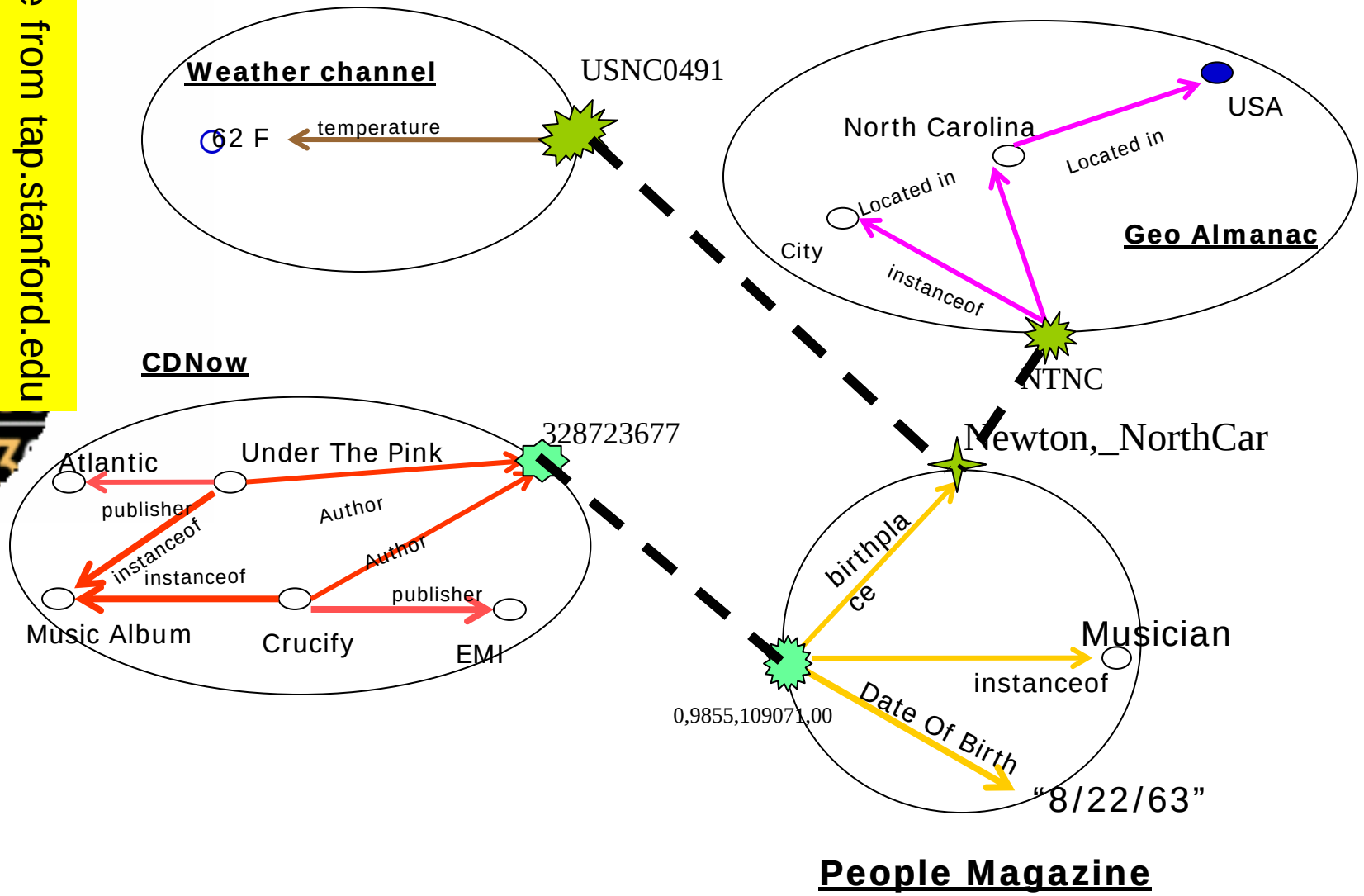


- Up to consumer to put these chunks together
- Situation analogous to pre-web hypertext systems and RDBMS today

This slide from tap.stanford.edu

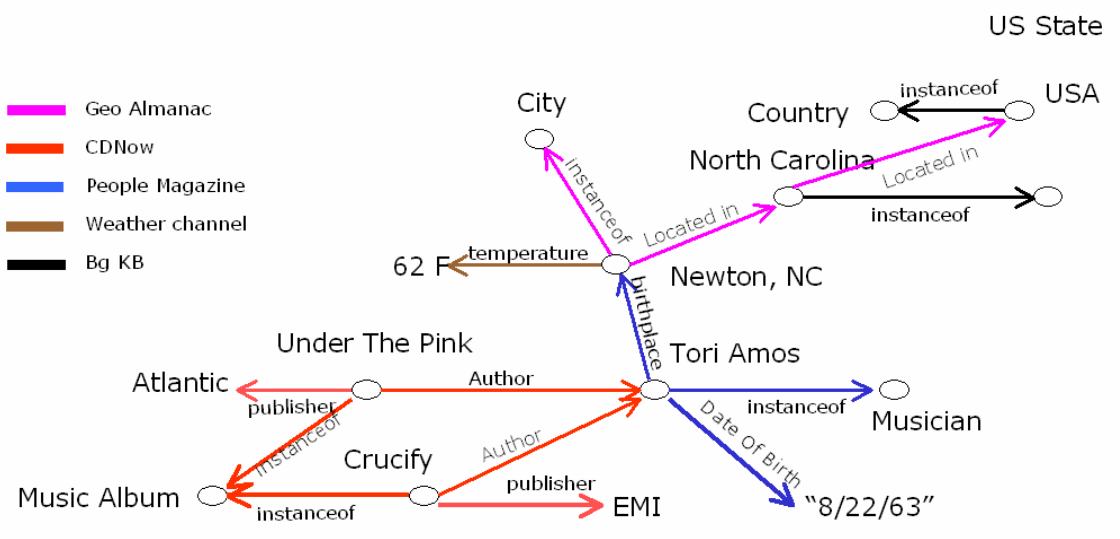
This slide from tap.stanford.edu

Semantic Web problem



TAP Goal

- Create a coherent semantic web from disparate chunks

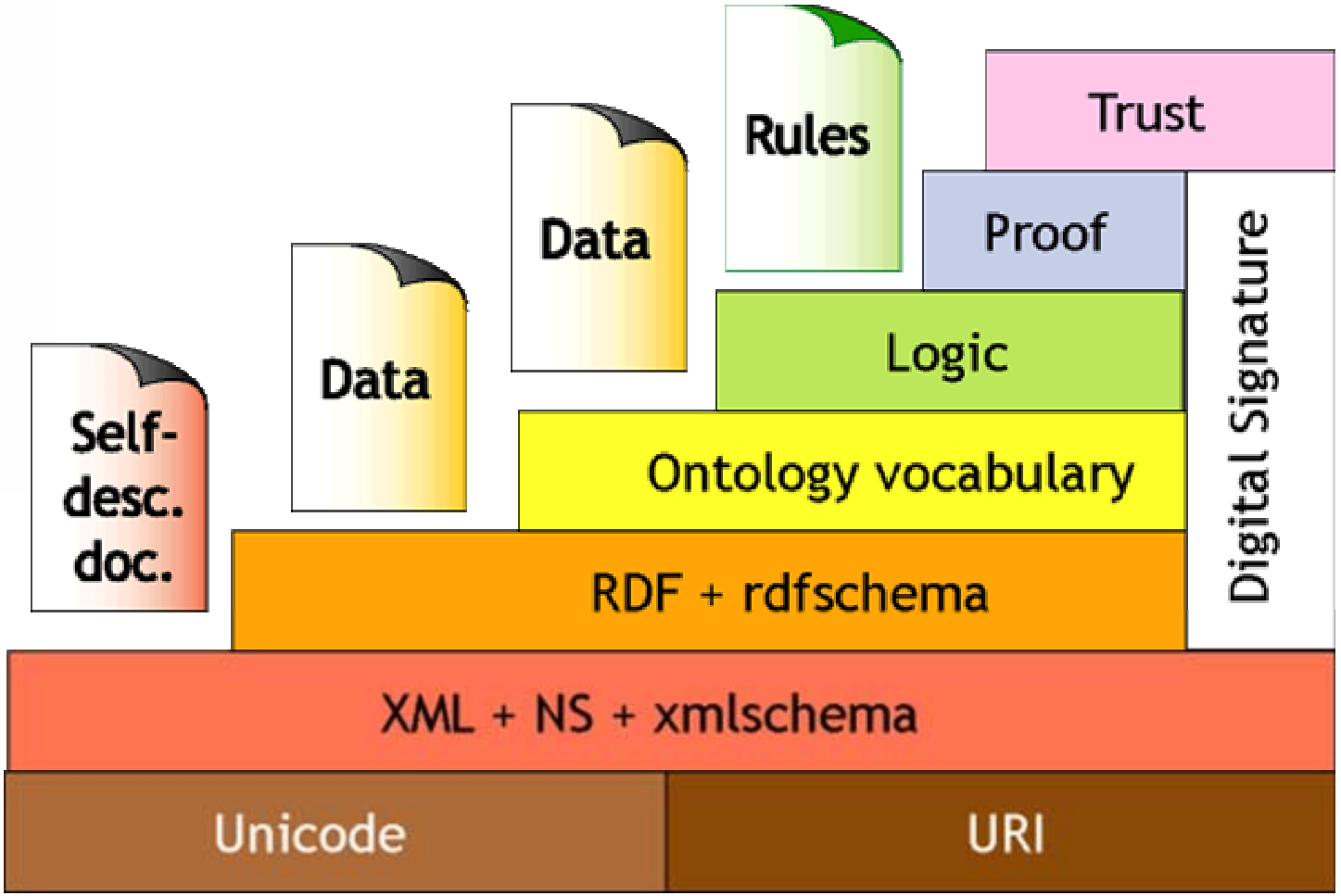


- Effectively make the web a giant distributed DB
- Why --- Bringing the Internet to programs

This slide from tap.stanford.edu

SW Layer Cake

-- Berners-Lee (99)
Swarts-Hendler (01)



Resource Description Framework

A knowledge representation format

- Encodes knowledge in sets of triples

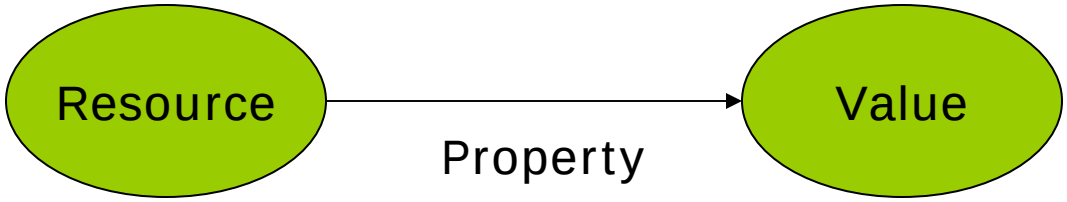
A document makes assertions that:

1. particular things (people, Web pages or whatever)
2. have properties (such as "is a sister of," "is the author of")
3. with certain values (another person, another Web page).

RDF Model

A model for representing named properties and property values

- models the equivalence relation
- Simply a triple of the form:



CS 5244 is offered in Fall 2004

RDF / XML: assertion interchange

- Simplified XML Syntax for RDF
  - Encodes RDF as machine parsable XML
  - Verbose, not really readable by humans
    - Note: counter to what one of XML's primary motivations.
- RDF and XML are complementary:
  - XML only gives structure (validating with a DTD)
  - RDF adds to XML the ability to encode simple propositions

RDF Schema – Basis for ontology

- RDF with XML: encode assertions
  - Still need to be able to exchange and reason on the data
  - To build the necessary ontology, RDF Schema was designed to be a **simple data typing** model for RDF

I need some F4 tickets

What's an F4? What's a ticket?

RDF Schema Core classes,properties,constraints

- rdfs:Resource
  rdfs:Property
  rdfs:Class
- rdf:Type
  rdfs:subClassOf
  rdfs:PropertyOf
- rdfs:ConstraintResource
  rdfs:ConstraintProperty
  rdfs:range
  rdfs:domain

Out of scope for today

RDF Schemas

- The first three most important concepts in RDF datatyping schema:
  - Resource (rdfs:Resource)
    - are objects that are uniquely identified by an URI
      - Note: URI not URL. Question: What is a URI?
  - Property (rdf:Property)
    - express the relationships of values associated with resources
  - Class (rdfs:Class)
    - are resources denoting a set of resources

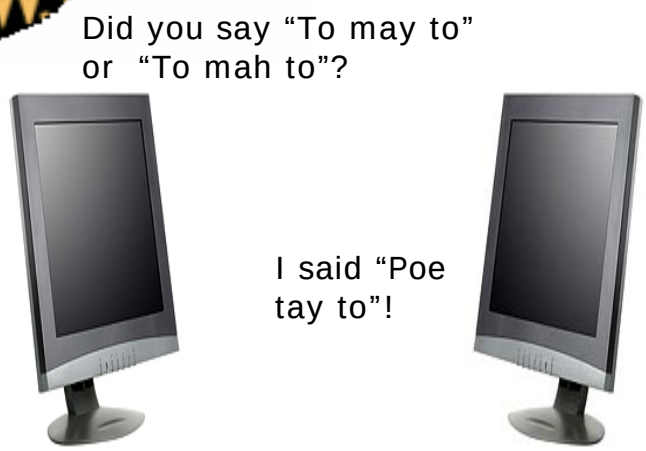
RDF schema example

```
Book rdf:type rdfs:Class .
:bookTitle rdf:type rdf:Property .
:bookTitle rdfs:domain :Book .
:bookTitle rdfs:range rdfs:Literal .
:MyBook rdf:type :Book .
:MyBook :bookTitle “My Book”
```

1. There’s a type of resource called “Book”
2. There a type of property called “BookTitle”
3. “BookTitle”s are a property of “Book”s
4. … and they can take a literal string value
5. MyBook is a type of Book
6. MyBook’s title is “My Book”

What about incompatible schemas? SW's Answer: OWL

- RDF Schema is fine if one person/organization is authoring all of SW
  - Inconsistencies among different authors
- OWL strengthens RDF Schema with some 30 additional interchange properties



owl:samePropertyAs = A and B same across schemas
owl:inverseOf, A is inverse property of B
owl:TransitiveProperty, allows transitivity
(e.g. A → B, B → C, then A → C)

References

- SW ontology development information (DAML):
  - http://derpi.tuwien.ac.at/~andrei/daml.htm
- Introduction to RDF Schema
  - http://www.dlib.org/dlib/may98/miller/05miller.html
- RDF and RDF schema
  - www.wastl.net/download/slides/rdf_overview.pdf
- OWL
  - http://www.w3.org/TR/owl-ref/

To think about…

- What are XML namespaces and how do they figure into the RDF syntax?
- Minimalist architecture makes the web scalable, will it make the SW workable?
- SW is not (yet fully) standardized
  - Help everyone out and see what you can contribute!
- What's your prediction when the SW will "arrive"?