Ontologies, the semantic web and RDF

Lecture 2 at Masaryk University

Nils Pharo

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Content

- bibliographic languages
- document languages and work languages
- subject languages - the LIS way
- ontologies - the CS way

An ontology is an explicit specification of a conceptualization.

What does this mean?

OSLO AND AKERSHUS UNIVERSITY COLLEGE OF APPLIED SCIENCES

Bibliographic languages

Document languages

– "A document is a particular space-time embodiment of information: a document language describes and provides access to this embodiment." (p. 107)

Work languages

– "describe information entities, their intellectual (as opposed to physical) attributes, and relationships among them." (p. 87)

Svenonius (2000). The Intellectual Foundation of Information Organization

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Document languages

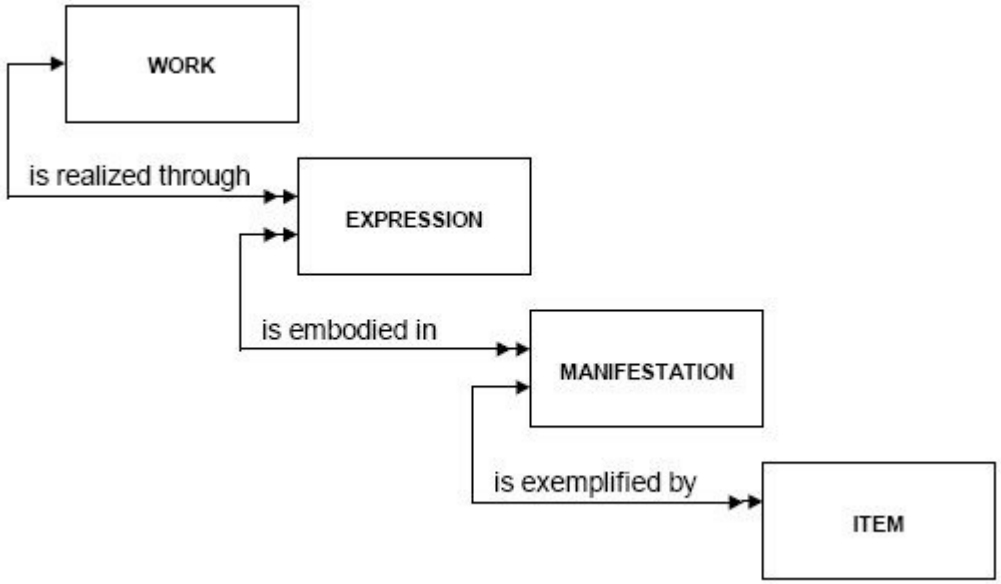
- Production language
- Carrier language
- Location language

Purpose of document languages

- for describing the material embodiment - the manifestation of the work
  - its physical and carrier attributes
  - its publication attributes
  - its external access attributes

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

FRBR



OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Work languages

- Author languages
- Title languages
- Edition languages
- Subject languages
  - Classification languages
  - Index languages

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Author, title and edition languages

- controlled and uncontrolled vocabularies
- normalized name forms for authority files
- uncontrolled names for descriptive cataloging

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Subject languages

- organized with respect to semantic strongness
  - "free keywords"
  - keyword lists
  - taxonomies
  - thesauri
  - faceted classification

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Keyword lists

- the most primitive form of controlled vocabulary
  - biology
  - horses
  - primates
  - psychology
  - wars

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Taxonomy

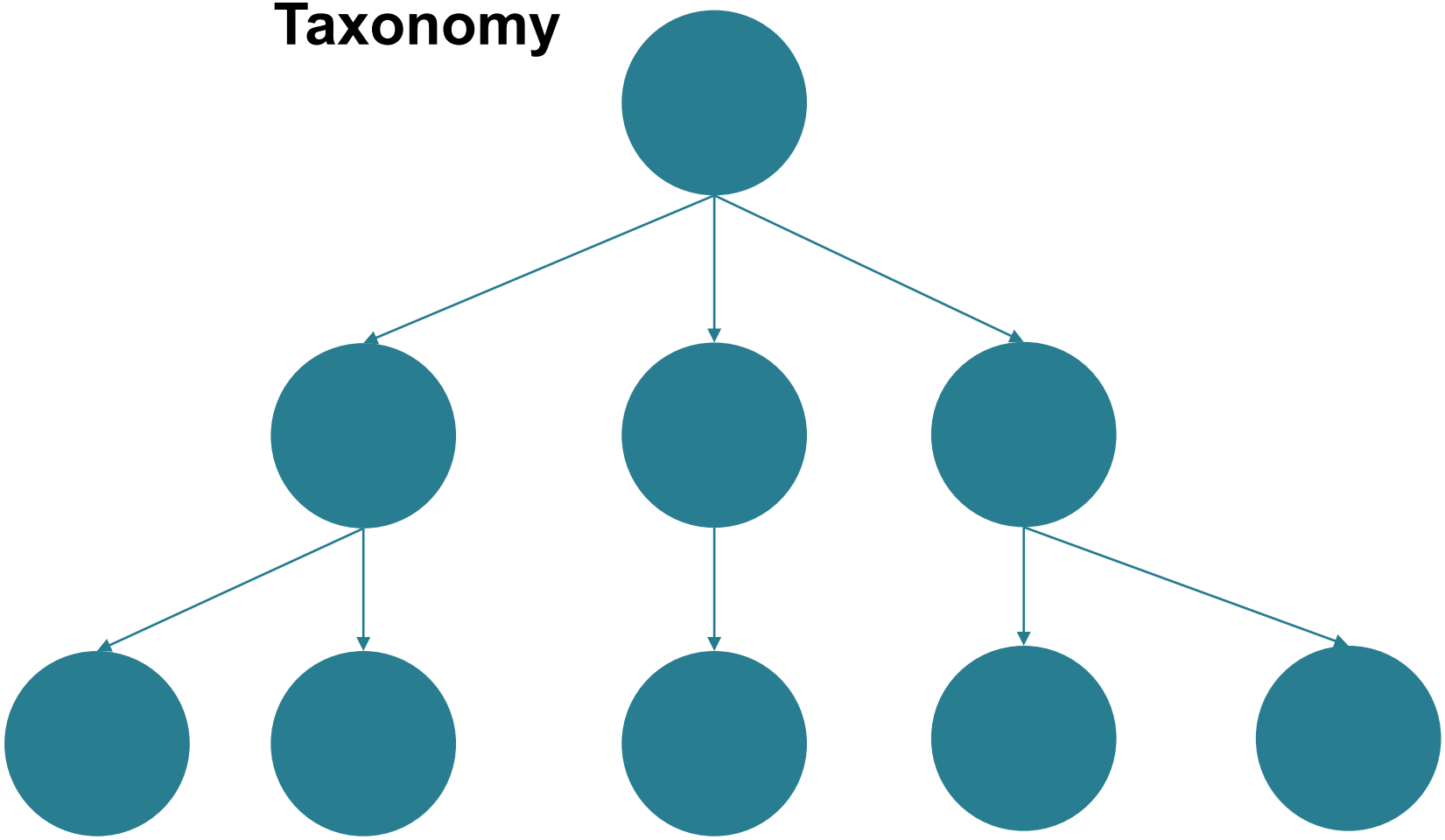
- hierarchical keyword list where terms are organized as subtypes/supertypes

- Animals
  - Cats
  - Dogs
  - Horses

- Food
  - Bread
  - Butter
  - Vegetables

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Taxonomy



OSLO AND AKERSHUS UNIVERSITY COLLEGE OF APPLIED SCIENCES

Thesauri

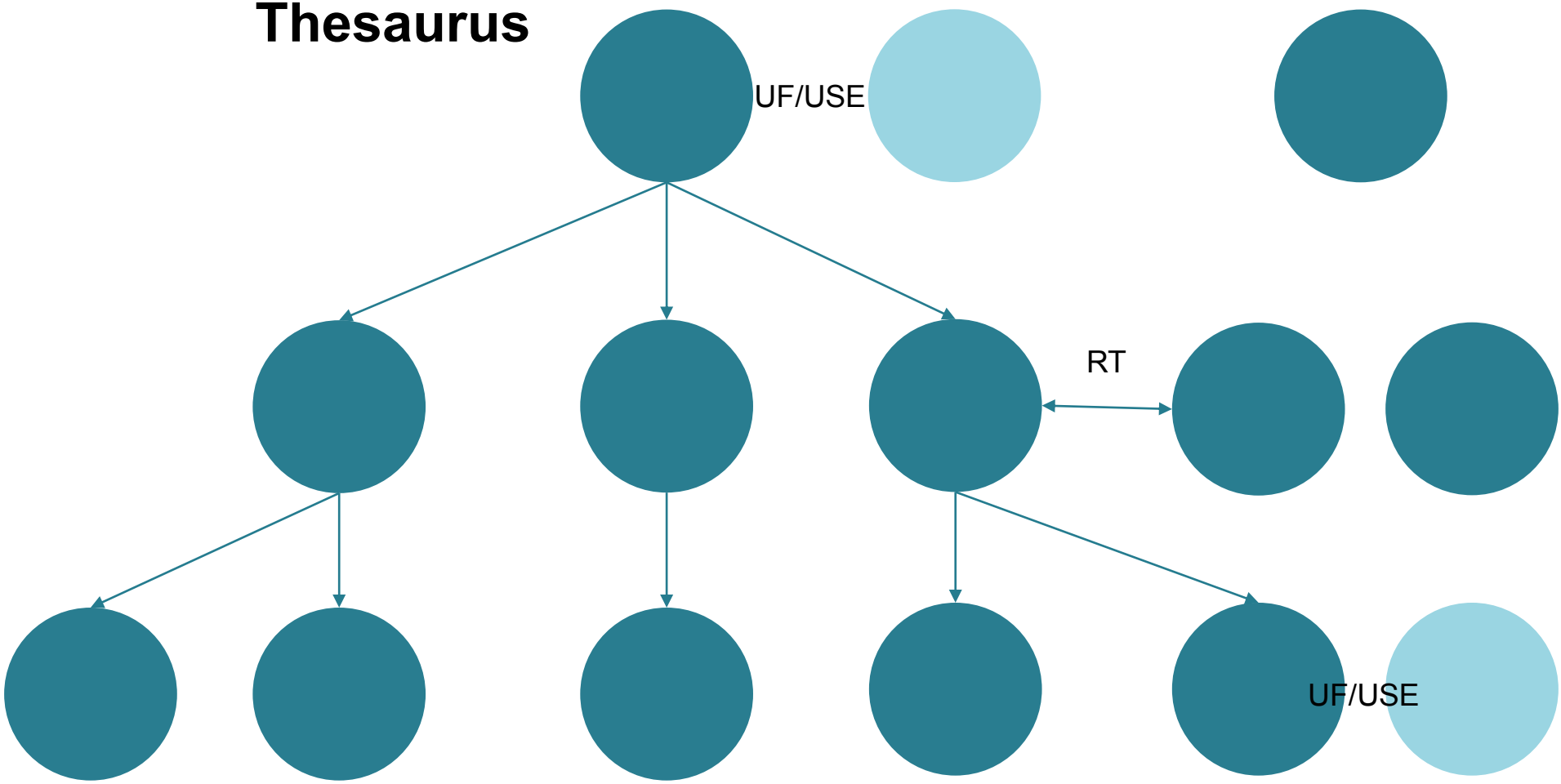
CHEFS

- UF Cooks
- BT Catering personnel
- RT Food preparation

Aitchison, Gilchrist & Bawden (2000). Thesaurus construction and use: a practical manual. (p. 164)

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Thesaurus



OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Thesaurus construction rules

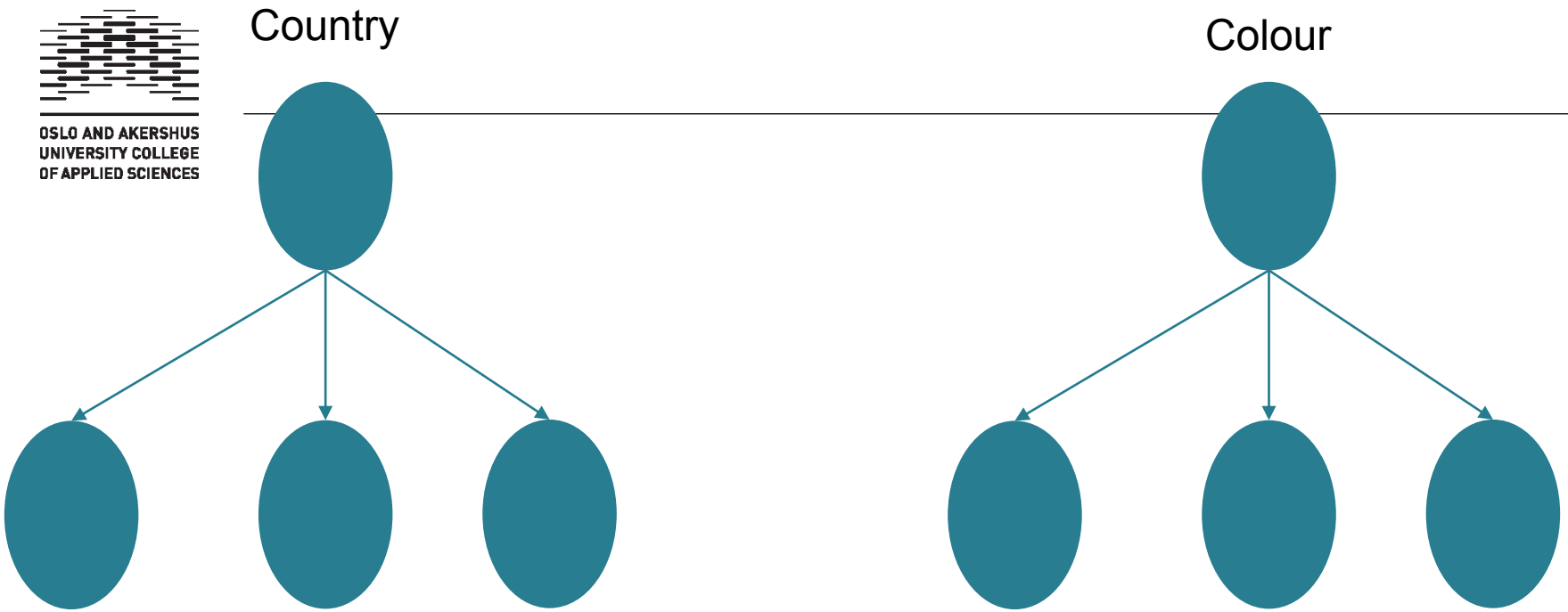
- three types of relationships: hierarchical, equivalence and associative
- scope notes are used to provide definitions, restrict use, clarify content of term etc
- standards (ISO 2788 and ISO 5964) that prescribe the implementation of the relationships
- recommendations for what associative relationships to be realised

OSLO AND AKERSHUS UNIVERSITY COLLEGE OF APPLIED SCIENCES

Facetted classification

- Wine
  - by region
    - France
    - Germany
    - Italy
  - by colour
    - Red
    - White
    - Rose
  - by price
    - less than 100 NOK
    - between 100 and 200 NOK
    - more than 200 NOK

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES



OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Charateristics of facetted classification

- no "standard"
- guidelines, e.g. Spiteri (1998). A simplified model for facet analysis
- Ranganathan's Colon classification

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Enter ontologies

- Original definition (from philosophy): the branch of metaphysics dealing with the nature of being.

- Adapted by computer scientists to facilitate artificial intelligence: "An ontology is an explicit specification of a conceptualization.[...] For AI systems, what "exists" is that which can be represented. When the knowledge of a domain is represented in a declarative formalism, the set of objects that can be represented is called the universe of discourse." Gruber (1993). Toward Principles for the Design of Ontologies Used for Knowledge Sharing

Ontologies

- Eye
  - Synonym:
    - Orbital part of face
    - Orbital region
  - Part:
    - Upper eyelid
    - Lower eyelid

From: Digital Anatomist Foundational Model of Anatomy ontology

Domain ontologies and top ontologies

- Domain ontologies models a specific domain, e.g. the human body, libraries, bread etc.
- Top ontologies describe concepts that are sharable across many domains.

Ontology components

- instances (individuals, entities, things)
- classes (types)
- properties (attributes, characteristics)
- relationships (relations)
- rules and constraints

Classes and instances

- instances represent concrete individuals or objects
- classes represent the collection of objects or individuals
- classes may contain other classes

Nils Pharo is an instance of the class person

Properties

used to denote aspects of the classes, instances and relationships

- <has a name> Fido
- <has a value> $1000
- <has an age> 5 years
- <has breed> boxer
- <has domain>
- <value range>
- <has an identifier>
- etc.

Relationships

specify how objects are related to other objects in the ontology, the most prominent being

- hierarchical superclass/subclass-relationships
  - dog <is a subclass of> mammal
- part-relationships
  - tail <is part of> dog

However, other forms of hierarchical relationships as well as relationships representing associative relations can be implemented

The structure of ontologies

- a hierarchical basic structure
- properties can be inherited
  - from superclass to subclass
  - Mammal *has hair*
  - Dog *has hair*
- instances can belong to multiple classes
  - Fido <is a> dog
  - Fido <is a> brown thing

Rules and constraints

to secure against illogical inferences, specify cardinality, and clarify the kinds of statements than can be used for specific classes, e.g.

- an animal cannot be both a carnivore and a herbivore
- an employee needs to be at a certain level of authencity to get access to high-security information
- a month cannot have more than 31 days

Purpose of ontologies

- model (a restricted part of) the world
- to make it possible for computers to infer things about the world
  - needs to be explicit!
- Open world assumption; a statement may be true irrespective of whether or not it is known to be true (Wikipedia article)

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Ontology questions

- how to model?
- are instances part of the ontology?
- what is the appropriate level of abstraction?
- what do the classes/instances refer to?

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Ontologies compared to old knowledge organization systems

- more flexible
- less standardized

- solution for merging and sharing
- solutions for identity

Ontology standard languages

- Topic maps (ISO)
- RDF/OWL (W3C)

Group work

– Create a simple ontology on a topic of your own choice!

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Content

- the semantic web
- why do we need it?
- the RDF standard
- interoperability recapitulated
- the data silo problem
- linked data

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

The semantic web

"The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF)." (http://www.w3.org/2001/sw/)

Challenges for the World Wide Web

The current Web is challenged on several areas

– too much noise
– internal systems with bad communication capabilities (data silos)
– large costs of communication

The noise problem

– more sophisticated IR-systems might help a bit
– needs more sophisticated mark-up

e.g.

Cantilever bridge

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Internal systems problem

- difficult to share data
- difficult to compare data
- difficult to reuse data

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Costs of communication

End-users need to

- collect
- interpret
- compare
- connect

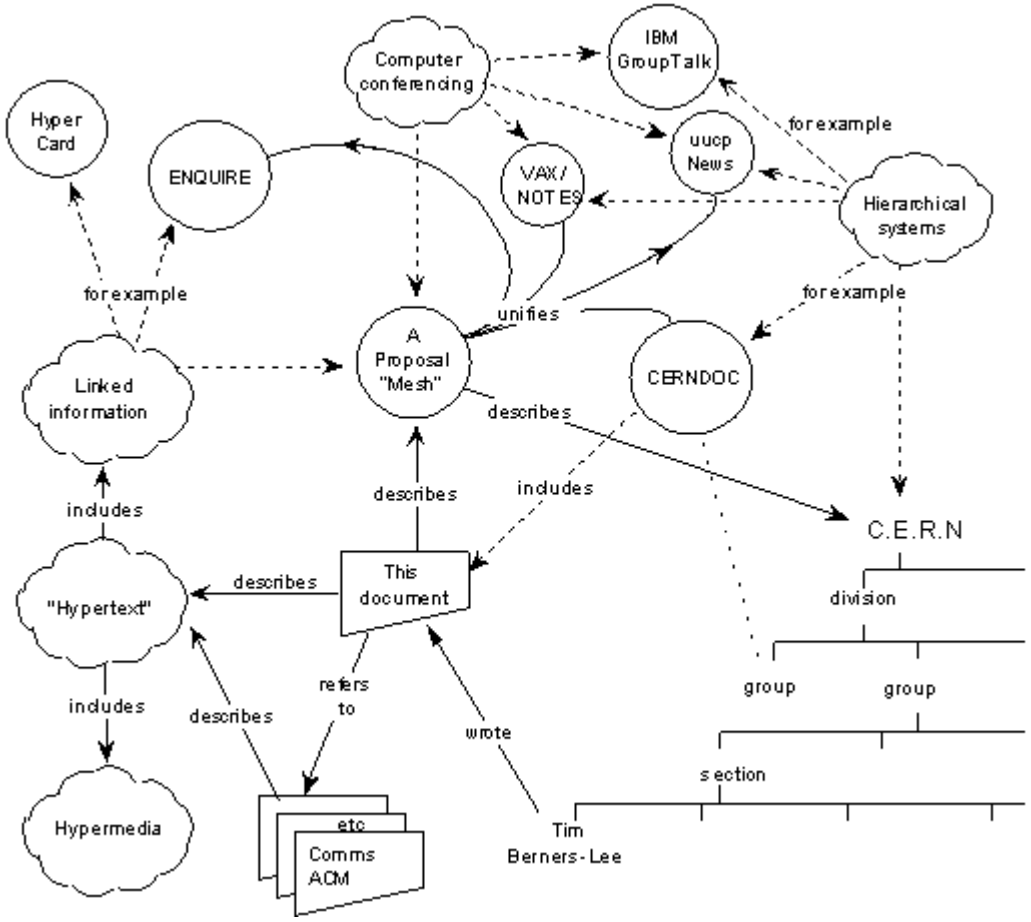
information themselves

Is the semantic web the solution?

- partial solution
- problem/domain dependent

The web for computer applications

- the SW is not intended to be interpreted by humans
- data semantically marked up and structured to be processed by intelligent agents
- SW is an extension of the Web
- SW is a web of data



– From [Tim Berners-Lee's 1989] proposal

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Ontologies, modelling domains for the semantic web

"An ontology is an explicit representation of a conceptualization" (Gruber, 1992)

- concepts and concretes modelled as classes (man)
- relationships (<is a> mammal)
- properties (<has name>
- constraint rules

to provide a: "shared and common understanding of a domain that can be communicated between people and application systems" Towards the semantic web, 2003

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Technologies for developing the semantic Web

W3C standard technologies

- XML
- RDF
- RDF Schema (RDFS)
- OWL (Web Ontology Language)

XML

– XML represents internal metadata to the item/document

Example: <author>Tim Berners-Lee</author>

RDF

- W3C standard (recommendation, 22.02.99)
- http://www.w3.org/RDF
- semantic Web - http://www.w3.org/2001/sw/
- tool for embedding metadata in digital documents

RDF describes

- things (subjects)
- properties (predicate)
- values (objects)

preferably identified by URIs

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Domain independent

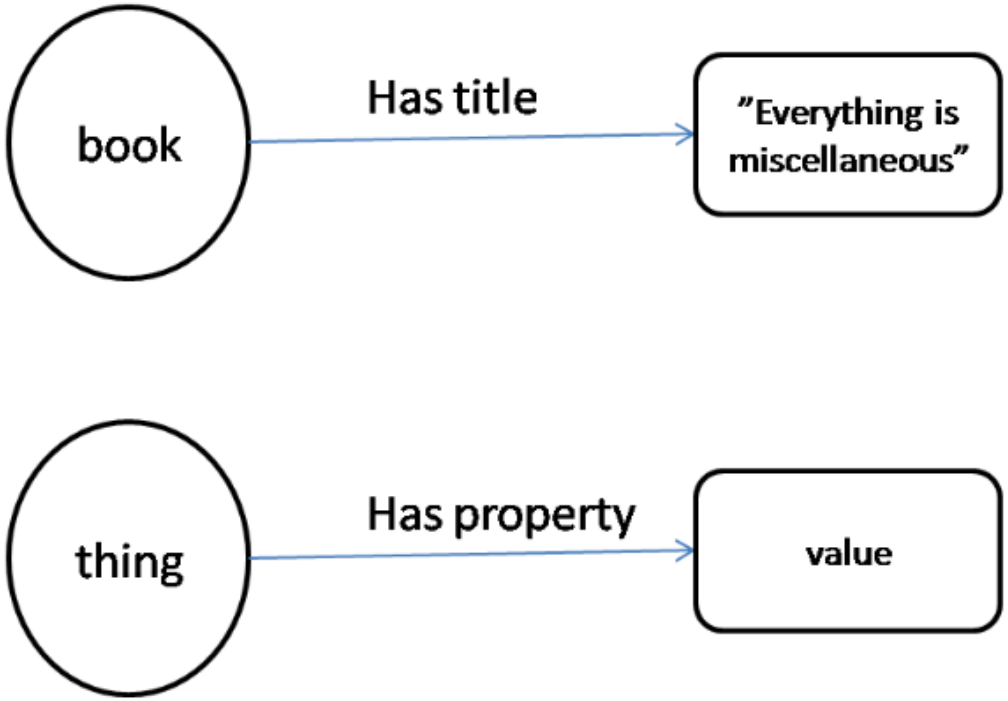
- RDF is a domain independent data model
- RDF describes triples representing things that have properties with values

<thing>Nils Pharo <property> is a teacher of <value> Digital knowledge organization

Relational database model

Books		
Isbn	**Author**	**Title**
1-932394-20-6	Thomas B. Passin	Explorer's guide to the Semantic Web
0-262-19433-3	Elaine Svenonius	The intellectual foundation of information organization
0-8050-8043-8	David Weinberger	Everything is miscellaneous

RDF model


book
Has title
"Everything is miscellaneous"
thing
Has property
value

OSLO AND AKERSHUS UNIVERSITY COLLEGE OF APPLIED SCIENCES

RDF example

http://www.jbi.hio.no/bibin/dig_korg/sem_web.htm has a **creator** whose value is **Nils Pharo**

In RDF/XML syntax:

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-
syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/">
<rdf:Description
rdf:about="http://www.jbi.hio.no/bibin/dig_korg/sem_web.htm
">
<dc:Creator>Nils Pharo</dc:Creator>
</rdf:Description> </rdf:RDF>
```

OSLO AND AKERSHUS UNIVERSITY COLLEGE OF APPLIED SCIENCES

Bibliographic RDF example

```
<?xml version="1.0"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.0/"
xmlns:rev="http://en.wikipedia.org/wiki/Book_review">
<rdf:Description
rdf:about="http://www.oreilly.com/catalog/9780596527341/">
<dc:title>Information Architecture for the World Wide
Web</dc:title>
<dc:creator>Peter Morville</dc:creator>
<dc:creator>Louis Rosenfeld</dc:creator>
<dc:publisher>O'Reilly</dc:publisher>
<dc:date>2006</dc:date>
<dc:language>en</dc:language>
<rev:review rdf:resource="http://www.digital-
web.com/articles/information_architecture_for_the_world_wide_web_3r
d_edition/"/>
</rdf:Description>
</rdf:RDF>
```

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Bibliographic example 2

```
<rdf:Description rdf:about="http://www.digital-
web.com/articles/information_architecture_for_the_world_wid
e_web_3rd_edition/">
<dc:title> A review of Information Architecture for the
World Wide Web, 3rd edition</dc:title>
<dc:creator>Lee McKusick</dc:creator>
<dc:publisher>PenLUG</dc:publisher>
<dc:date rdf:datatype=
"http://www.w3.org/2001/XMLSchema#date">2006-23-
12</dc:date>
<rev:review_of
rdf:resource="http://www.oreilly.com/catalog/9780596527341/
"/>
</rdf:Description>
</rdf:RDF>
```

OSLO AND AKERSHUS UNIVERSITY COLLEGE OF APPLIED SCIENCES

Notation 3

A simpler syntax for human readiability

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/">
<rdf:Description
rdf:about="http://www.jbi.hio.no/bibin/dig_korg/sem_web.htm">
<dc:Creator>Nils Pharo</dc:Creator>
<dc:Subject>Semantic Web</dc:Subject>
</rdf:Description> </rdf:RDF>
```

equals:

```
<@prefix dc: <http://purl.org/dc/elements/1.1/>.
<@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
<http://www.jbi.hio.no/bibin/dig_korg/sem_web.htm> dc:Creator
"Nils Pharo" ;
dc:Subject "Semantic Web" .
```

RDF describes instances

The rdf:type property can be used to state that a resource is an instance of a class

- RDF schema is a simple ontology language
- OWL is a full ontology language

RDF schema

- RDF schema is used for defining RDF terminologies
- RDF schema is a type system for RDF
- RDF schema makes semantic information machine-accessible
- RDF schema is a simple ontology language

Example: the statement "Nils Pharo is a teacher of Digital knowledge organization" can be used to deduce that "Nils Pharo is a member of the academic staff" and that "Nils Pharo is involved with Digital knowledge organization"

key components: class, subclass relations, property, subproperty relations, domain and range constraints

OSLO AND AKERSHUS UNIVERSITY COLLEGE OF APPLIED SCIENCES

RDFS example

```
<?xml version="1.0"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xml:base="http://example.org/staff#">
<rdf:Description rdf:ID=”Employee">
<rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-
schema#Class"/>
</rdf:Description>
<rdf:Description rdf:ID=”Teacher">
<rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-
schema#Class"/>
<rdfs:subClassOf rdf:resource="#Employee"/>
</rdf:Description>
</rdf:RDF>
```

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

RDFS example 2

```
<?xml version="1.0"?>

<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xml:base="http://example.org/staff#">
<rdfs:Class rdf:ID=”Employee" />

<rdfs:Class rdf:ID=”Teacher">
  <rdfs:subClassOf rdf:resource="#Employee"/>
</rdfs:Class>
<rdf:Property rdf:ID=”employedBy”/>
</rdf:RDF>
```

RDFS example 3 (Notation 3-format)

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema# >.
@prefix rdf: < http://www.w3.org/1999/02/22-rdf-syntax-ns#
>.
<http://example.org/staff#Employee> a rdfs:Class .
<http://example.org/staff#Teacher> a rdfs:Class ;
rdfs:subClassOf <http://example.org/staff#employee>
```

OWL -Web ontology language

- funded on DAML+OIL
- OWL is a richer ontology language than RDF schema
- 3 versions supporting different levels of complexity; Full, DL, and Lite

can be used to specify that: "academic staff members must teach at least one course" or "every book must have a title"

OWL elements

- OWL uses RDF, RDF schema and its own terminology to define ontologies:
- Web Ontology Language: OWL by Grigoris Antoniou and Frank van Harmelen which includes an example of an OWL-defined ontology
- W3 org's OWL guide

Assignment

Model the hierarchical parts of the ontology you constructed previously with the use of RDF and RDFS