Ontologies and Folksonomies

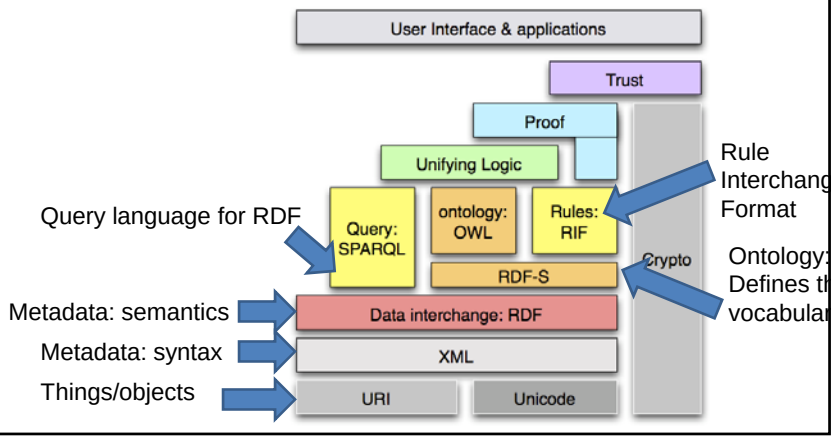
Social Computing Class 2009
Julita Vassileva
University of Saskatchewan

SOFIA: http://picasaweb.google.ca/julitav/DropBox?authkey=Gv1sRgCK-3v5mB5a38Dg#5399921463445091762

How to organize the Web so that we can find stuff?

- The Semantic Web: (T. Berners-Lee et al., 2001)

... is an extension of the current web in which information is given well-defined **meaning**, better enabling computers and people to **work in co-operation**.



Approaches to organize knowledge

- The Philosophical / Artificial Intelligence approach - Ontologies (from Wikipedia):
  - **Ontology** (from the Greek ὄν, genitive ὄντος: *of being* (neuter participle of εἶναι: *to be*) and -λογία, -logia: *science*, *study*, *theory*) is the philosophical study of the nature of being, existence or reality in general, as well as of the basic categories of being and their relations.
  - In computer science and information science, an **ontology** is a formal representation of a set of concepts within a domain and the relationships between those concepts. It is used to reason about the properties of that domain, and may be used to define the domain. An ontology is a "formal, explicit specification of a shared conceptualization".[1] An ontology provides a shared vocabulary, which can be used to model a domain
- The Social Web approach
  - Tags and Folksonomies

Ontologies

- Taxonomies
  - Layers, layers, layers of metadata
  - Various metadata standards

•Let's play a standards Acronym trivia:
XCBF , XKMS, SAML, XACML, WSML

  - WordNet
- Ontologies - inter-related entities, structures
  - Since the mid-1970s, researchers in the field of artificial intelligence have recognized that capturing knowledge is the key to building large and powerful AI systems. AI researchers argued that they could create new ontologies as computational models that enable certain kinds of automated reasoning.

Metadata

HTML provides formatting

```
<H1>Social Computing</H1>
    <UL>
        <LI>Teacher Name
        <LI>Student Name
    </UL>
```

XML provides syntax

```
<course>
    <title>Social Computing</title>
    <teacher> Teacher Name</teacher>
    <student>Student Name</student>
</course>
```

RDF provides metadata about web resources

```
<rdf:Description rdf:about="#York">
  <tel>6086758</tel>
</rdf:Description>
```

RDF Schema adds vocabulary for RDF.

Organizes the vocabulary in typed hierarchy

- Class, subClassOf, type
- Property, subPropertyOf
- Domain, range

RDF (resource description framework)

RDF is based on the idea of identifying things using Web identifiers (called *Uniform Resource Identifiers*, or *URIs*), and describing resources in terms of simple properties and property values.

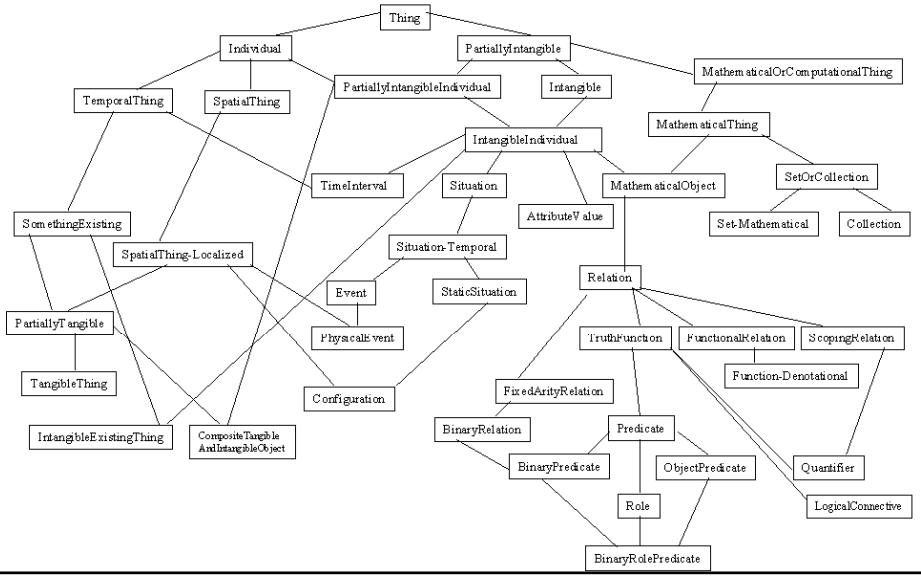
```
<?xml version="1.0"?>
<rdf:RDF
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#">
   <contact:Person rdf:about="http://www.w3.org/People/EM/contact#me">
      <contact:fullName>Eric Miller</contact:fullName>
      <contact:mailbox rdf:resource="mailto:em@w3.org"/>
       <contact:personalTitle>Dr.</contact:personalTitle>
   </contact:Person>
</rdf:RDF>
```

RDF Identifies:

individuals, e.g., Eric Miller,
identified by http://www.w3.org/People/EM/contact#me
kinds of things, e.g., Person,
identified by http://www.w3.org/2000/10/swap/pim/contact#Person
properties of those things, e.g., mailbox,
identified by http://www.w3.org/2000/10/swap/pim/contact#mailbox
values of those properties, e.g. mailto:
em@w3.org as the value of the mailbox property

RDF Tutorial (30 min)

Ontologies



Ontology components and languages

Common components of ontologies include (Wikipedia)

- Individuals: instances or objects (the basic or "ground level" objects)
- Classes: sets, collections, concepts, types of objects, or kinds of things.[10]
- Attributes: aspects, properties, features, characteristics, or parameters that objects (and classes) can have
- Relations: ways in which classes and individuals can be related to one another
- Function terms: complex structures formed from certain relations that can be used in place of an individual term in a statement
- Restrictions: formally stated descriptions of what must be true in order for some assertion to be accepted as input
- Rules: statements in the form of an if-then (antecedent-consequent) sentence that describe the logical inferences that can be drawn from an assertion in a particular form
- Axioms: assertions (including rules) in a logical form that together comprise the overall theory that the ontology describes in its domain of application.
- Events: the changing of attributes or relations

Ontology languages

An ontology language is a formal language used to encode the ontology.

- IDEF5 is a software engineering method to develop and maintain usable, accurate, domain ontologies.
- KIF is a syntax for first-order logic that is based on S-expressions.
- Rule Interchange Format (RIF) and F-Logic combine ontologies and rules.
- OWL is a language for making ontological statements, developed as a follow-on from RDF and RDFS, as well as earlier ontology language projects including OIL, DAMLand DAML+OIL. OWL is intended to be used over the World Wide Web, and all its elements (classes, properties and individuals) are defined as RDF resources, and identified by URIs.

Example: FOAF ontology for social relationships

- http://www.foaf-project.org/
- Classes:
  | Agent | Document | Group | Image | OnlineAccount | OnlineChatAccount | OnlineEcommerceAccount |OnlineGamingAccount | Organization | Person | PersonalProfileDocument | Project |
- Properties:
  | accountName | accountServiceHomepage | aimChatID | based_near | birthday | currentProject | depiction | depicts| dnaChecksum | family_name | firstName | fundedBy | geekcode | gender | givenname | holdsAccount | homepage | icqChatID |img | interest | isPrimaryTopicOf | jabberID | knows | logo | made | maker | mbox | mbox_sha1sum | member | membershipClass| msnChatID | myersBriggs | name | nick | openid | page | pastProject | phone | plan | primaryTopic | publications |schoolHomepage | sha1 | surname | theme | thumbnail | tipjar | title | topic | topic_interest | weblog | workInfoHomepage |workplaceHomepage | yahooChatID |

Attempts to harvest ontology power

- Ontology Search engine Swoogle: http://swoogle.umbc.edu http://ebiquity.umbc.edu/project/html/id/53/ (description)
- But it works only for semantically annotated sites
- Humans annotating content – currently the most common approach
- But how to annotate dynamic service content efficiently?
  - Semantic Deep Web crawlers - crawl repeatedly, constructing deep data signature for docs and services, then frequency distribution analyses and clustering … active area of research…

How do YOU organize

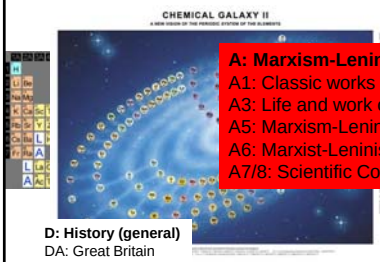
- Your kitchen cupboards?
- Your clothes in the closet?

- The files on your computer?
- Your digital photos?

"The solution to overabundance of data is more data"
David Weinberger

Principle limitations of ontologies

- An ontology always reflects a particular viewpoint, purpose or constraint (of its creator)
  - E.g. library catalogues optimize book shelves
    - physical location of books in library (one book in just one category)
    - of cataloguing

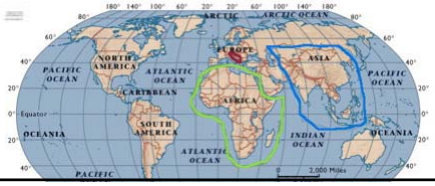


A: Marxism-Leninism
A1: Classic works of Marxism-Leninism
A3: Life and work of C.Marx, F.Engels, V.I.Lenin
A5: Marxism-Leninism Philosophy
A6: Marxist-Leninist Political Economics
A7/8: Scientific Communism

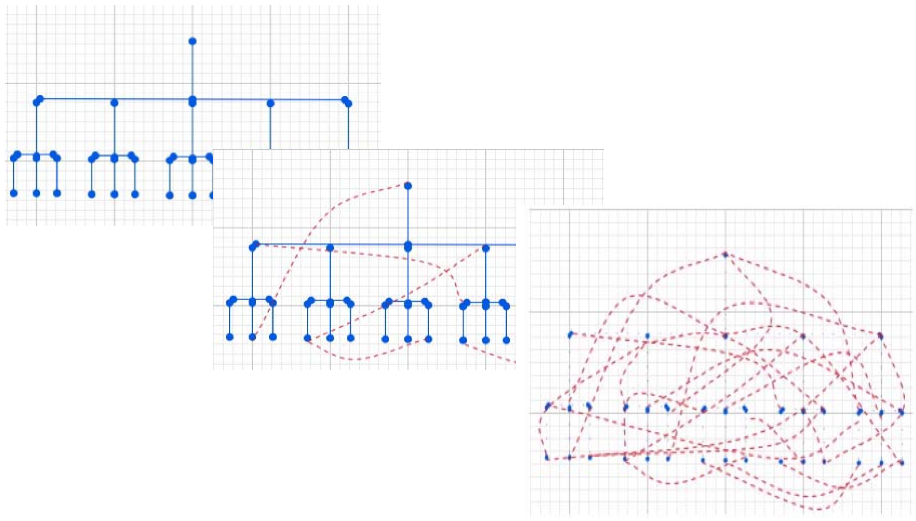
Dewey, 200: Religion
210 Natural theology
220 Bible
230 Christian theology
240 Christian moral & devotional theology
250 Christian orders & local church
260 Christian social theology
270 Christian church history
280 Christian sects & denominations
290 Other religions

D: History (general)
DA: Great Britain
DB: Austria
DC: France
DD: Germany
DE: Mediterranean
DF: Greece
DG: Italy
DH: Low Countries
DJ: Netherlands

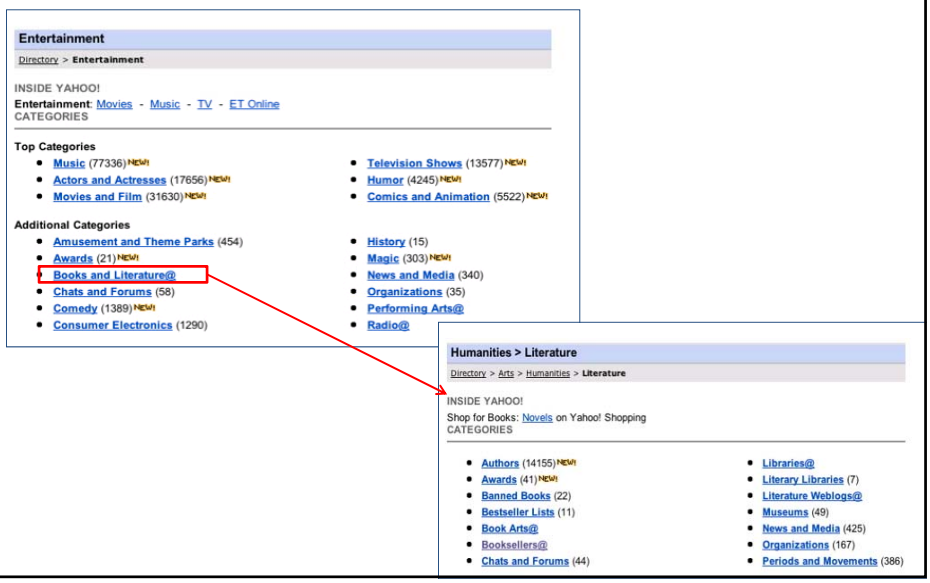
DP: Iberian Peninsula
DQ: Switzerland
DR: Balkan Peninsula
DS: Asia
DT: Africa
DU: Oceania
DX: Gypsies



Hierarchies and non-hierarchies



Example - Yahoo

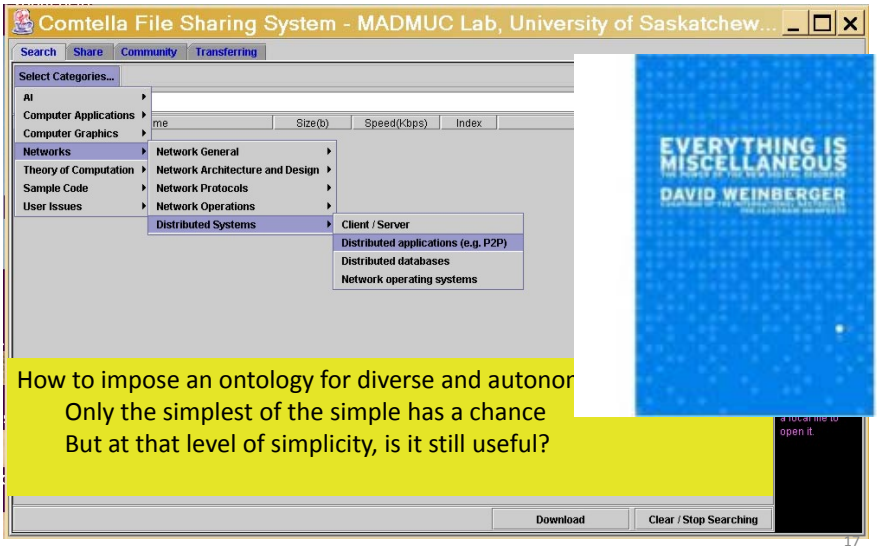


More problems

- Categorizing has aspects of
  - Mind reading (guessing how others will interpret)
  - Fortune telling (predicting the future)
- Categorizing leads to information loss
  - E.g. category of interest: “movies”, “films”, “cinema” ➔ are they all the same really?
  - “Smart people think differently”
- Different communities have implicit naming agreements: hard to find consensus
- Hard to agree upon the semantics of relationships
- Even if people agree formally, they may still interpret differently…

- Even simple hierarchies are hard to use

Use of taxonomy-based annotation



Summary

When does Ontological Classification work well?

- Small corpus
- Formal categories
- Stable entities
- Restricted entities
- Clear edges

Expert catalogers
Authoritative source of judgment
Coordinated users
Expert users

When does Ontological Classification NOT work well?

- Large corpus
- No formal categories
- Unstable entities
- Unrestricted entities
- No clear edges

Uncoordinated users
Amateur users
Naive catalogers
No Authority

Folksonomies

Folksonomies (web 2.0)

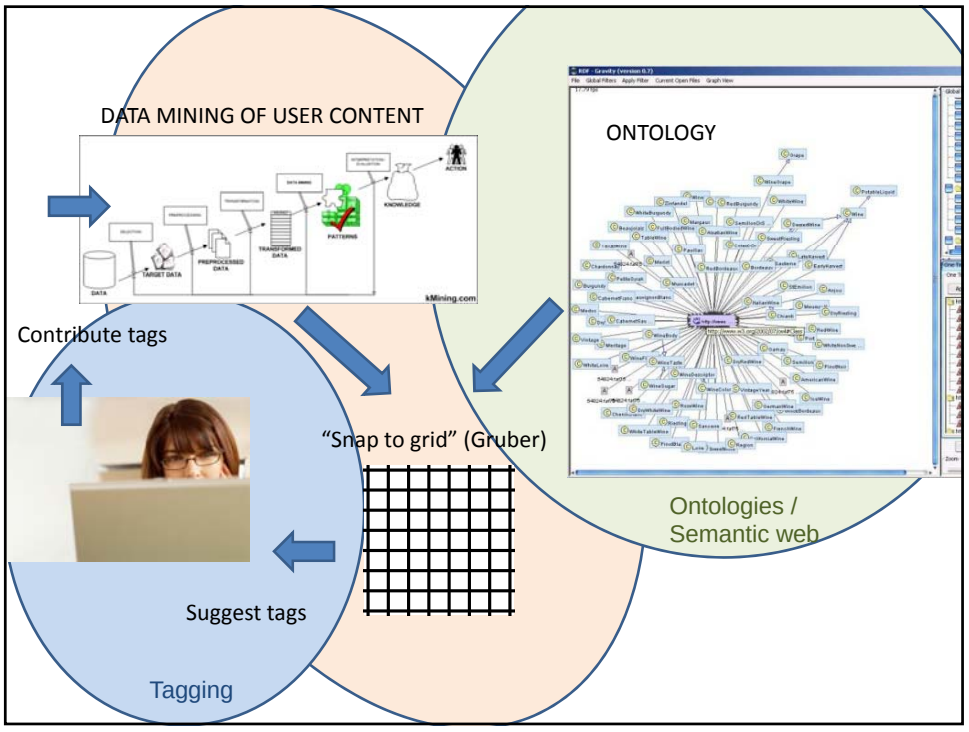
- Positives
  - Selfish users tag for themselves
  - Easy to add
  - Tags can express different semantic dimensions: content, context, pedagogical, learner-type, media type, --> similar to metadata
  - No standards (just help in avoiding misspellings)
  - Tag sharing leads to social quality control
  - Flat list of tags, font size indicates tag popularity
  - Tag “cloud” gives a summary of document (browsing)
  - Allows easy search by tag (instead of forming queries)

Folksonomies – problems:

- The machine does not know the semantics of the document without knowing how the tags relate to each other (i.e. an ontology of tags ☺)
  - Can't say how two documents are related or why they are similar (not qualitatively)

    → Hard to sequence a presentation from tagged materials

    → But for a "one-shot" retrieval tags are okay.



Features of these solutions

- User centered – respect user's autonomy
- Easy for the user – just like a folksonomy
- The AI happens in the background, the user is not aware of it
- Simplicity and ease of use preserved, advantages of ontology added

- http://www.bazaarblog.com/2007/10/28/everything-is-miscellaneous-as-told-by-video/