Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

Semantic Web – Introduction and Problem Statement

- Initial Concepts
- Initial Problems
- Key Perspectives
- The Role of XML
- Meta Data + RDF
- Representing Meaning
- Ontologies + OWL
- Evaluating Resources

Semantic Web: The Idea

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

Tim Berners-Lee, James Hendler, Ora Lassila: “The Semantic Web“.

Scientific American, May 2001

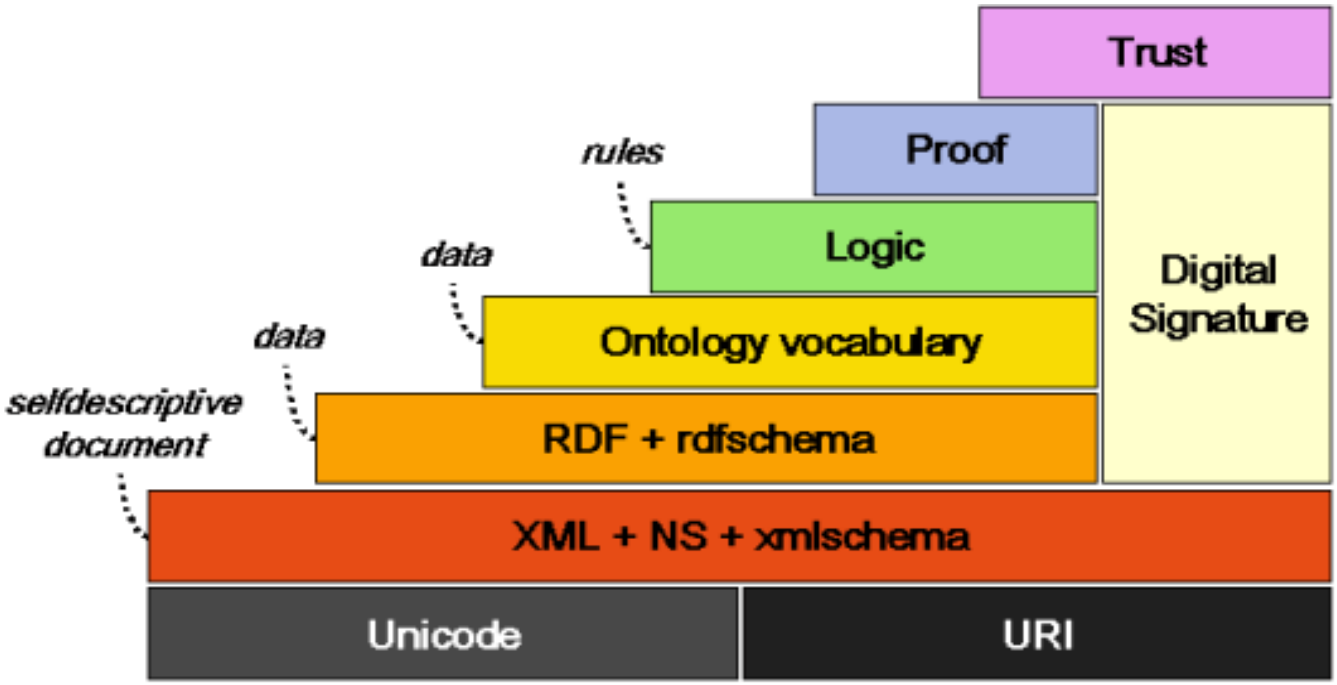
Objectives

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

- Bring machine processable structure to the bulk of Web information
- Provide a layer of meaningful meta information along with Web offers to identify their semantics
- Provide semantic rules to the community to digest Web concurrency and allow for conclusions
- Offer ways to learn about the reputation of a resource

Semantic Web Layers

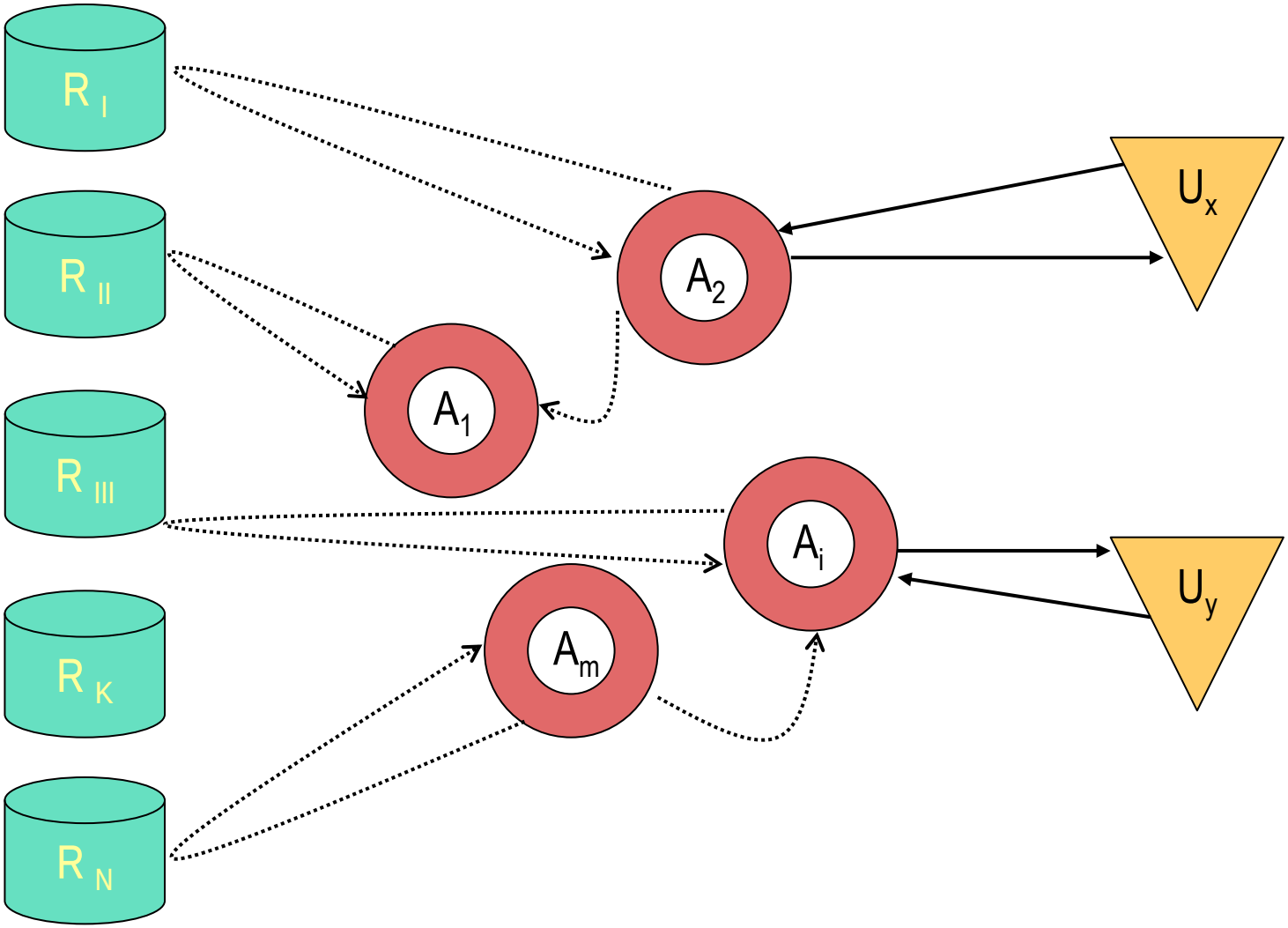
Thomas Schmidt
schmidt@informatik.
haw-hamburg.de



Source: http://www.w3.org/2001/12/semweb-fin/w3csw

Operational Concept

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de



Resources Agents Users

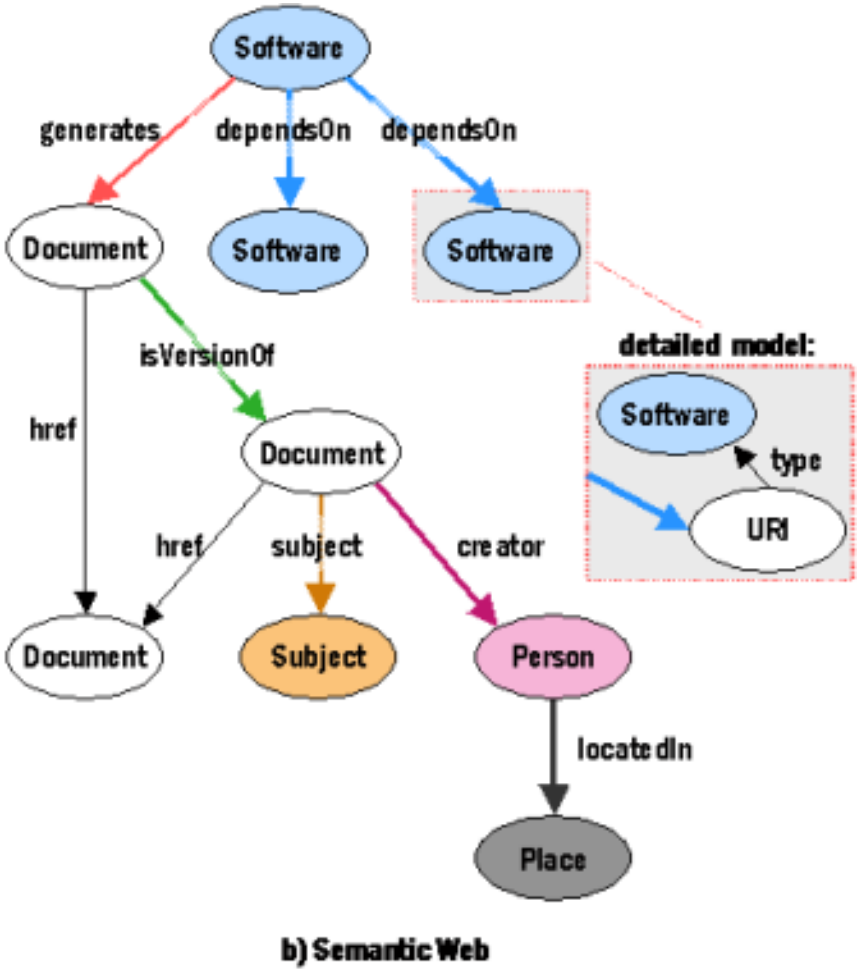
Resources, URIs & Links

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

Goal: Understand resources and their relations

Resources: Anything addressable by a URI. Extend resources to carry a ‘type’ attribute

Links: Relating resources. Extend links to carry ‘type’ attribute.



b) Semantic Web

Fundamental Problems

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

Heterogeneity: Systems, encoding, structures, languages/expressiveness', words, meanings, …

Anonymity: Almost all resources in the Web unknown to recipient

Context: Resources are meaningless without identification of context

Scale: Peer-to-Peer view has complexity n^2, with n = number of Internet resources

Visions & Expectations: partly naive, partly vague, …

Propagated Visions

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

- “The goal of the Semantic Web is to create a universal medium, which smoothly interconnects personal, commercial, scientific and cultural data in a machine-understandable fashion. ”
- “With the Semantic Web we can provide all kinds of automated services in different domains from future home and digital libraries to electronic business and health services.”
- “The Semantic Web taking over completely ones life, which is the ultimate goal.”
- … space for your profound statement …

Specific Problems

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

- Meaningful Buzz Word: Everybody makes up his own meaning (like ‘Artificial Intelligence’, ‘Chaos Theory’, …).
- The Hype: People without intricate understanding involved prior to proved results.
- Propagated Visions raise unrealistic expectations.
- Awkward Visions: Sacrificed stuff, people don’t want.
- Ridiculous Personality Cult about Tim Berners-Lee.

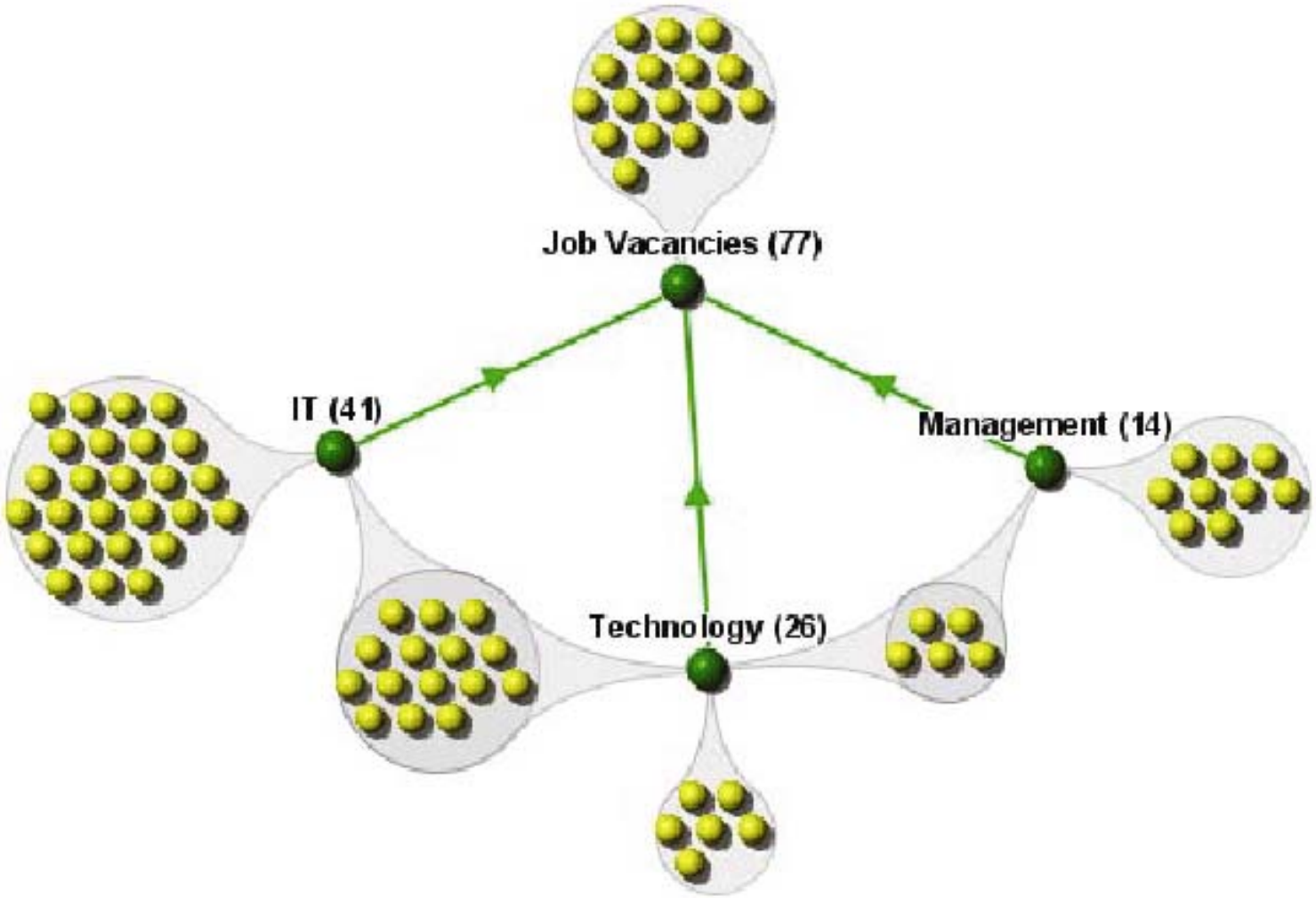
What May We Expect?

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

- Search & retrieval: Improved search machines and identification of information
- Data integration: High level tools for rapid/semi automated data source coupling
- High-level applications in specific fields: Knowledge management, eLearning, …
- …adaptive distributed systems,
- … original user interfaces for navigation & visualisations

Visualisation Example: Cluster Maps

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de



Data-centric Perspective

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

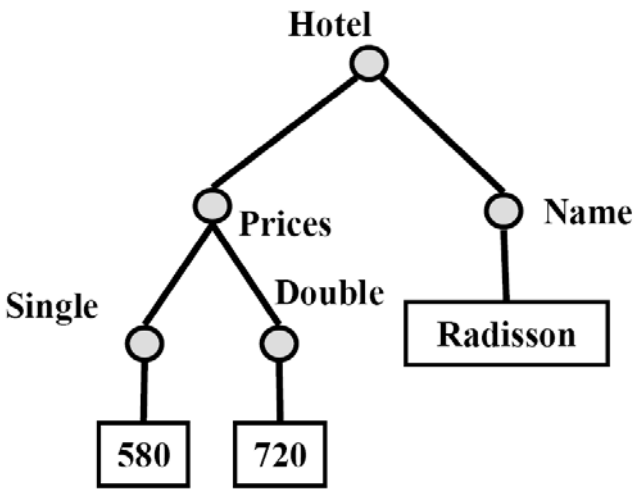
- Explore information, discover knowledge
  - Find and understand, what resources are about
  - Understand and digest content of a web resource
- Evaluate information
  - Estimate relevance and reputation
  - Judge on precision and correctness
- Integrate information
  - Combine data from different sources
  - Synchronize different data bases

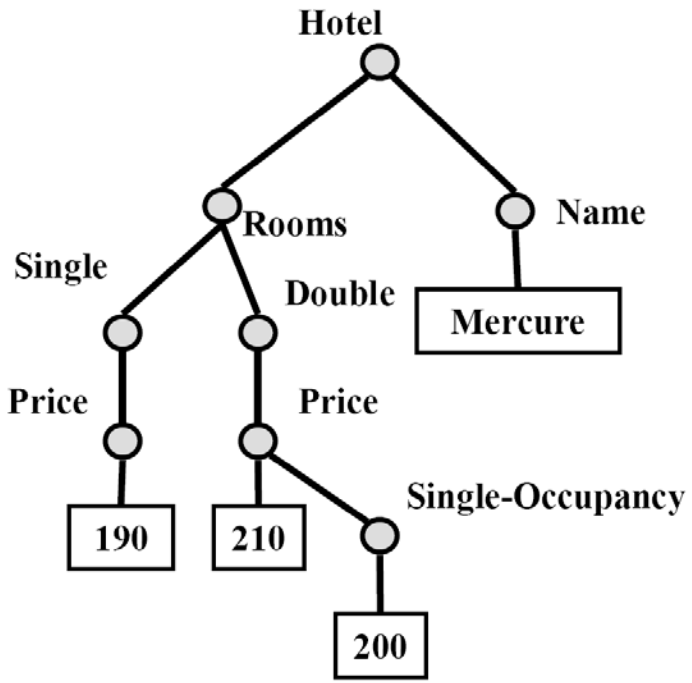
Structural Heterogeneity

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

Different data models with

- Naming / type / integrity conflicts
- Multilateral correspondences
- Missing / redundant / inconsistent data

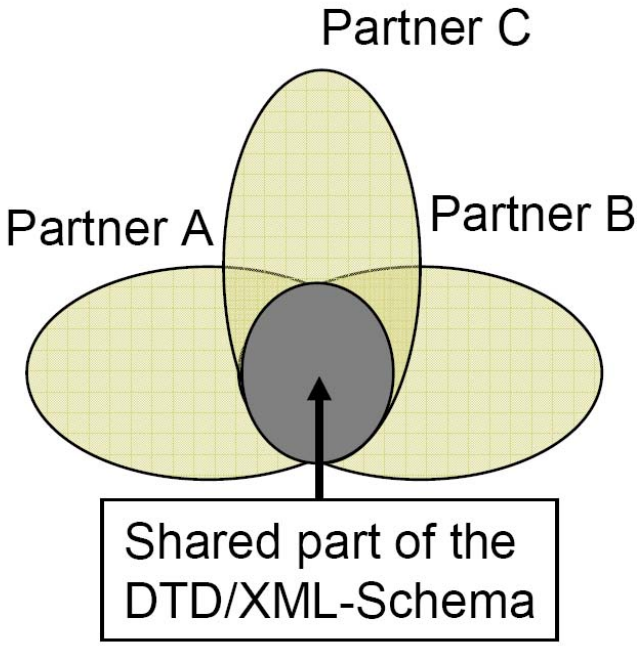




The Role of XML

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

- XML provides standards and transparency on syntax and encoding
- Plays role of a basic interoperability mechanism
- Data exchange formally solved with common DTD/Schema
- XML itself has no semantic definitional strength
- Meaningful only in communities with appointed agreements

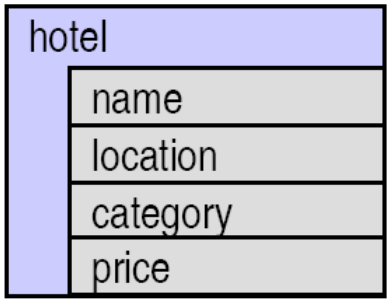


- Don’t forget: XHTML is mainly without structure!

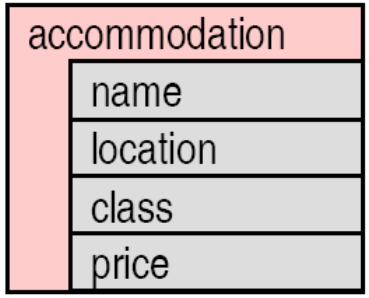
Semantic Heterogeneity

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

- Meta level discrepancies lead to diverging terms
- Data semantics may be divergent € vers. $



Name	Location	Category	Price
Radisson	Copenhagen, Denmark	Congress-Hotel	580
Mercure	Hamburg, Germany	Congress-Hotel	190
Ritz	London, England	Congress-Hotel	130
...	...	...	...



Name	Location	Class	Price
IBIS	Potsdam, Germany	Hotel	65
Meier	Berlin, Germany	Apartment	55
Schulz	Potsdam, Germany	Bungalow	60
...	...	...	...

How to Solve Heterogeneity Problems ?

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

1. Structural Heterogeneity
   - Comparing semantically corresponding data schema entities
   - Correlating semantically corresponding data attributes
   - Transforming correspondent data types (if possible)
   - Special Problem: Aggregation of multilateral correspondences
2. Semantic Heterogeneity
   - Detecting semantic correspondences in data

► Semantic is a key issue to solve data heterogeneity problems

Application-centric Perspective

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

Automated use of applicative resources (e.g. web services) require answers to:

- What does the application require ?
- How does it work ?
- How is it used ?

Application / community specific approaches:

- OWL-S: Semantic Markup for Web Services
- BPEL: Business Process Execution Language
- …

Focus of the DAML initiative (www.daml.org)

Meta Data (traditional)

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

- Provide a (formalized) description about resources and information
- Commonly organized as (property : values) maps
- Provide some structure on top of (arbitrary) data, subject to standardization
- Standards provide definitions on (structured) properties, occasionally a vocabulary of values

Dublin Core

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

- RFC2413 – Simple description scheme (http://dublincore.org)
- Initially minimal consensus from a working group of librarians

Content	Intellectual Property	Instantiation
Title	Creator	Date
Subject	Publisher	Format
Description	Contributor	Identifier
Type	Rights	Language
Source		
Relation		
Coverage		

Learning Object Metadata

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

- IEEE standard scheme for describing learning objects (LOs)
- Provides a defined, extensible vocabulary in 9 categories

1 General General information describing the learning object as a whole.	**2 Lifecycle** Documenting the history and the current state of the LO as well as its contributors.	**3 Meta Metadata** Groups information about the meta data instance itself.
4 Technical Technical requirements and characteristics.	**5 Educational** Allows for a list of educational and pedagogic characterizations.	**6 Rights** Intellectual property rights and conditions of use.
7 Relation A list of qualified descriptions of the relationship between this instance and other learning objects.	**8 Annotation** Comments on educational use of the learning object.	**9 Classification** Information about this learning object in relation to a particular classification system.

LOM - General

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

1.1	Identifier		A globally unique label that identifies this LO.
	1.1.1	Catalog	The name or designator of the identification or cataloging scheme for this entry. A namespace scheme.
	1.1.2	Entry	The value of the identifier within the identification or cataloging scheme for this entry. A namespace specific string.
1.2	Title		Name given to this LO.
1.3	Language		The primary human language or languages used within this LO to communicate to the intended user.
1.4	Description		A textual description of the content of this LO.
1.5	Keyword		A keyword or phrase describing the topic of this LO.
1.6	Coverage		The time, culture, geography or region to which this LO applies.
1.7	Structure		Underlying organizational structure of this LO.
1.8	Aggregation level		The functional granularity of this LO.

LOM - Educational

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

5.1	Interactivity Type	Predominant mode of learning supported by this LO.
5.2	Learning Resource Type	Specific kind of LO.
5.3	Interactivity Level	The degree of interactivity characterizing this LO.
5.4	Semantic Density	The degree of conciseness of a LO.
5.5	Intended End User Role	Principal user(s) for which this LO was designed.
5.6	Context	The principal environment within which the learning and use of this LO is intended to take place
5.7	Typical Age Range	Age of the typical intended user.
5.8	Difficulty	How hard it is to work with or through this LO for the typical intended target audience.
5.9	Typical Learning Time	Approximate or typical time it takes to work with or through this LO for the typical intended target audience.
5.10	Description	Comments on how this LO is to be used.
5.11	Language	The human language used by the typical intended user.

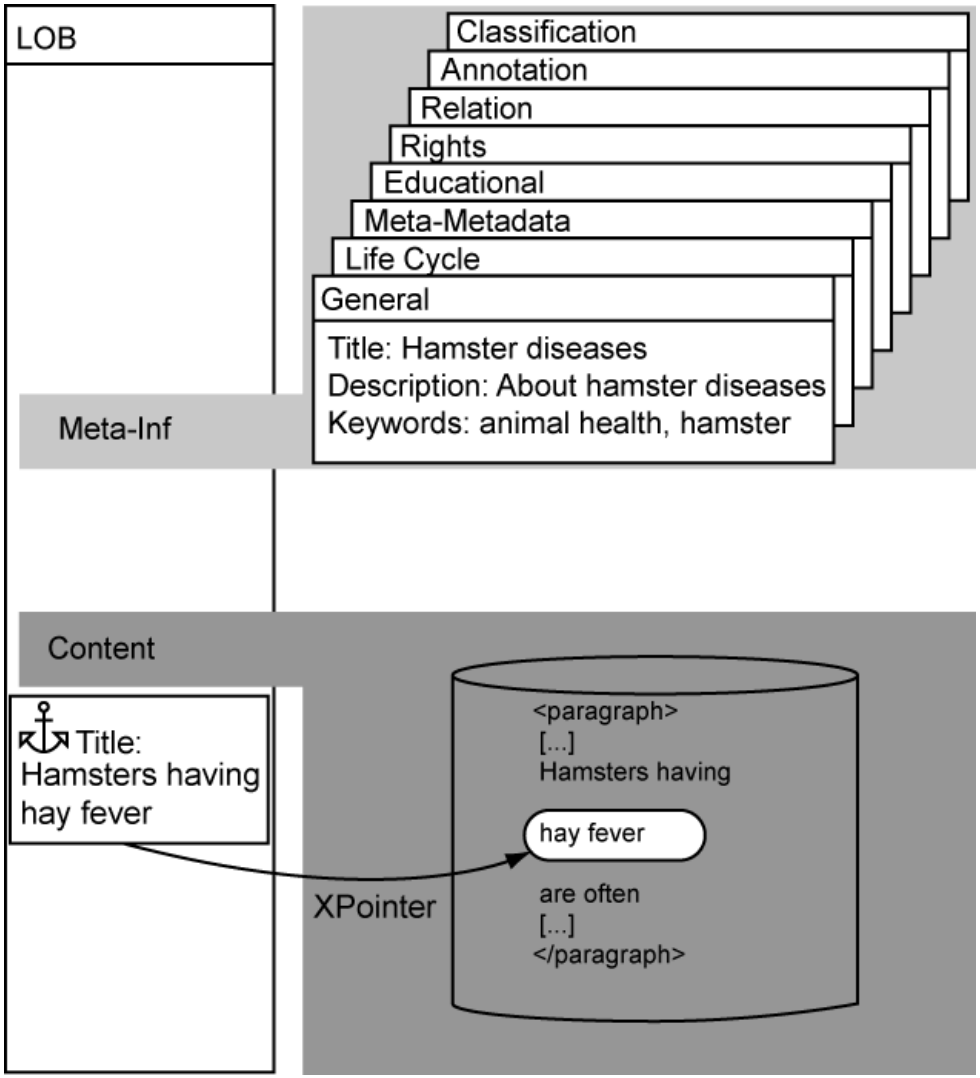
LOM - Relation

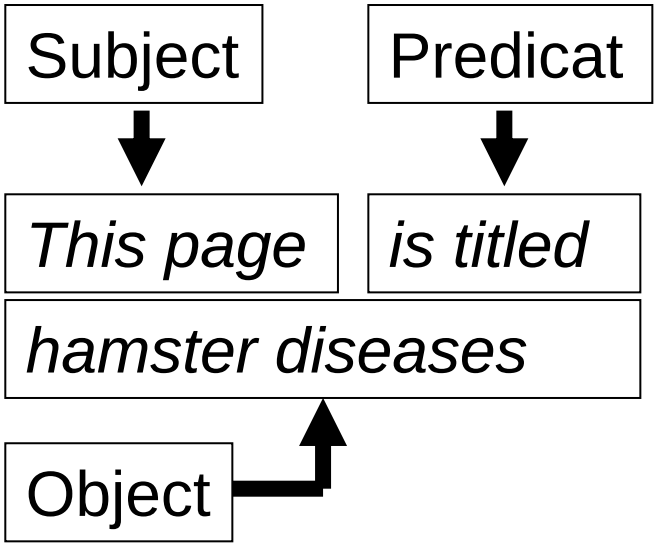
Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

7.1	Kind			Nature of relationship between this LO and the target LO identified by 7.2.
7.2	Resource			The target LO object that this relationship references.
	7.2.1	Identifier		A globally unique label that identifies the target LO.
		7.2.1.1	Catalog	The name or designator of the identification or cataloging scheme for this entry. A namespace scheme.
		7.2.1.2	Entry	The value of the identifier within the identification or cataloging scheme for this entry. A namespace specific string.
	7.2.2	Description		Description of the target LO.

Meta Data Extraction

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de





Resource Description Framework (RDF)

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

- Performs statements about resources
- Statements as triple
  - Subject + Predicat + Object
- Maps information directly and unambiguously to a decentralized model
- URIs used to name objects and concepts
- Graphical representation as semantic nets
- Syntax independent, but usually XML
- Two (syntactically differing) expressions equal if RDF models coincide

RDF Example

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

- Statements:
  - Subject: Resource
  - Predicat: Property
  - Object: Literal

“Hay fever handbook	has the author	Simon”

http://…

RDF Syntax

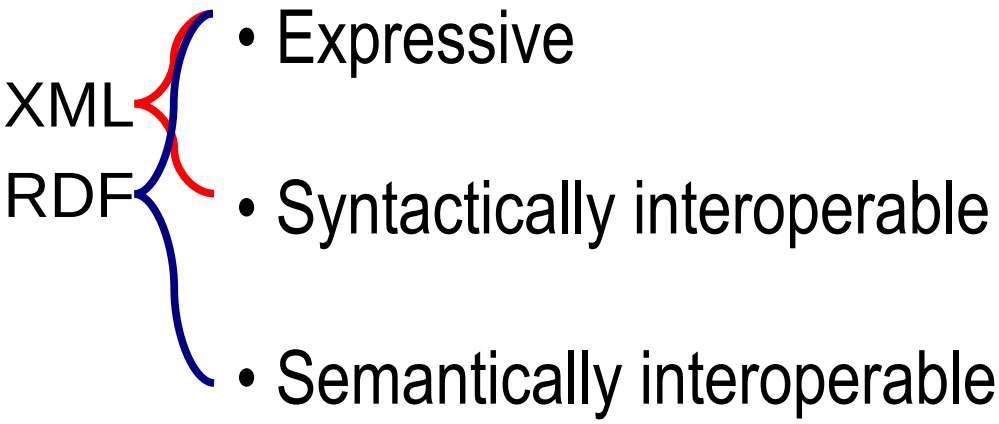
Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

```
<rdf:RDF>
  <rdf:Description about="http://... ">
    <dc:author xmlns:dc="http://purl.org/dc">
      Simon
    </dc:author>
  </rdf:Description>
</rdf:RDF>
```

- XML encoding
- Standard allows for abbreviation

Roles of XML and RDF in the Semantic Web

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de



Representing Meaning

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

- Words represent meaning
- Dictionaries define the meaning of words
- Problem: Many-to-many relation between words and meanings
  - There may be many words for one meaning
  - Words may have many, very distinct meanings
- Solution: Employment of controlled vocabularies
  - Pre-selected words used in pre-appointed meanings
- Approaches: use taxonomies and thesauri as present

Taxonomies

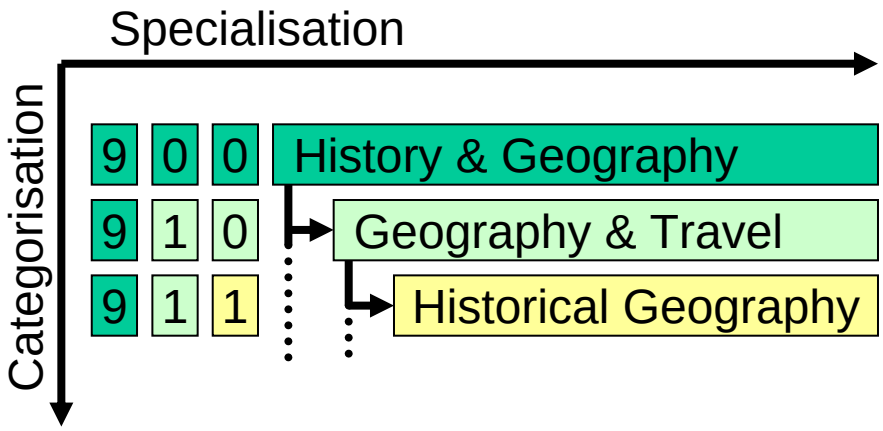
Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

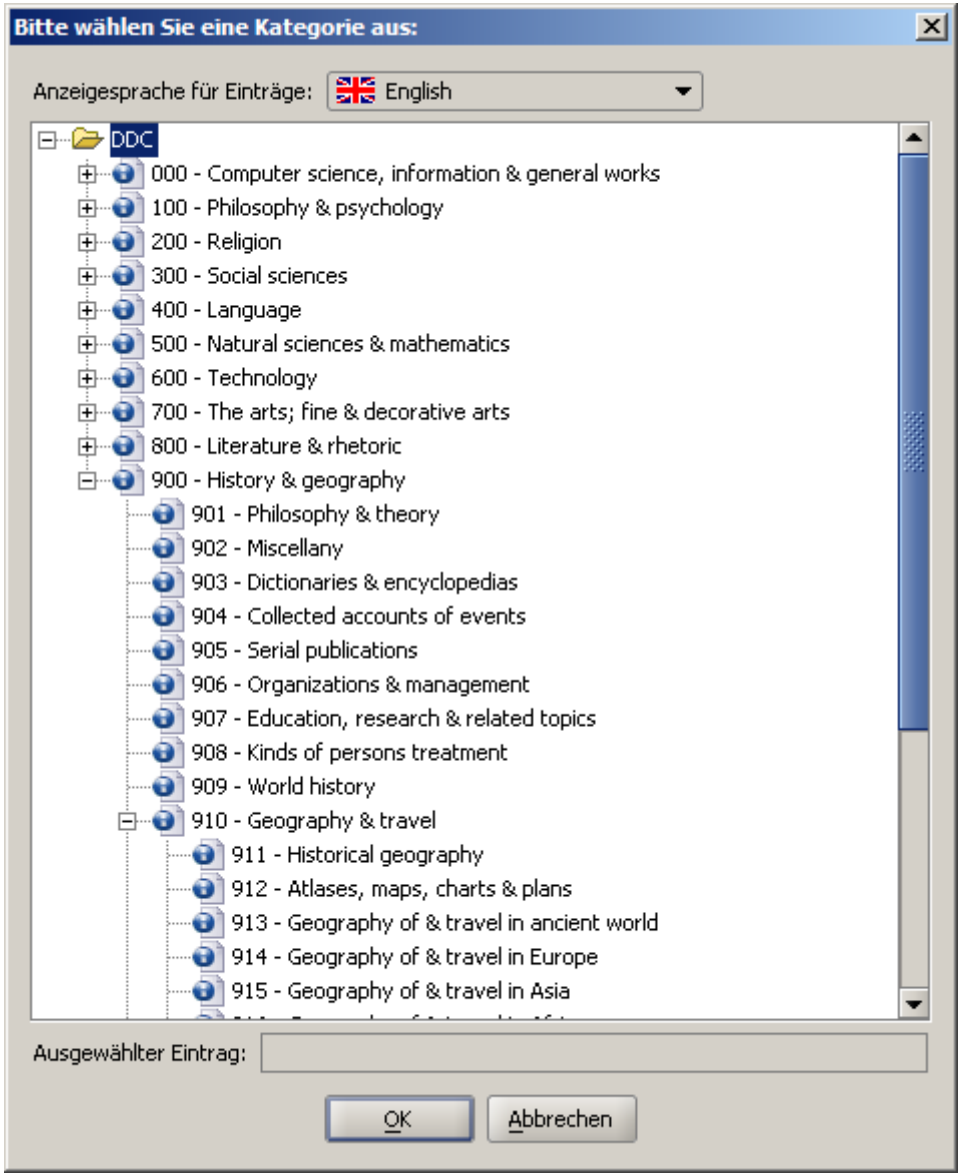
- A taxonomy is a hierarchy of notions, representing a systematic classification for a collection of entities
- Tree represents specialisation/generalisation
- Association represents an ‘is instance of’ relation
- Examples:
  - Linnaeus System (biology)
  - ACM Computer Science Index
  - Dewey Decimal Classification - DDC
  - North American Industry Classification System - NAICS
- Expressiveness: all categories in classified structure of a given context

Dewey Decimal Classification

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

- Classification of general knowledge within 10 main categories and 10 layers of hierarchy
- Designed by Melvil Dewey in 1873 owned by the OCLC since 1988
- Version 23 today, ≈ 33.000 entities
- Extensible and processable with minimal effort
- Some problems with structural evolution of knowledge and local dependencies (e.g. law)





Thesauri

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

- A thesaurus is a classification scheme for terms
- A taxonomy of terms (from a given language) + additional semantic relations:
  - Hierarchical (broader : narrower)
  - Equivalence (synonym : antonym)
  - Homographic
  - Associative
- Useful to extend a controlled vocabulary
- Example: Roget's Thesaurus (www.gutenberg.net)
- Expressiveness: vocabulary and its basic semantic relations for (parts of) a given language

Example: WordNet

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

http://www.cogsci.princeton.edu/~wn/

Results for "Hypernyms (this is a kind of...)" search of noun "buster"

5 senses of buster

…

buster, broncobuster -- (a person who breaks horses)

=> horseman, equestrian, horseback rider -- (a man skilled in equitation)
=> rider -- (a traveler who actively rides an animal (as a horse or camel))
=> traveler, traveller -- (a person who changes location)
=> person, individual, someone, somebody, mortal, human, soul a human being; "there was too much for one person to do")
=> organism, being --(a living thing that has (or can develop) the ability to act or function independently)
=> living thing, animate thing -- (a living (or once living) entity)
=> object, physical object a tangible and visible entity; an entity that can cast a shadow; ”it was full of rackets, balls and other objects")
=> entity -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))
=> causal agent, cause, causal agency -- (any entity that causes events to happen)
=> entity -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

TopicMaps

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

- Originally form ISO standard (HyTime 19744)
- Goals
  - Intelligent information retrieval and subsequent processing
  - Accessing a semantic network of knowledge
  - Putting hypertext in semantic relations

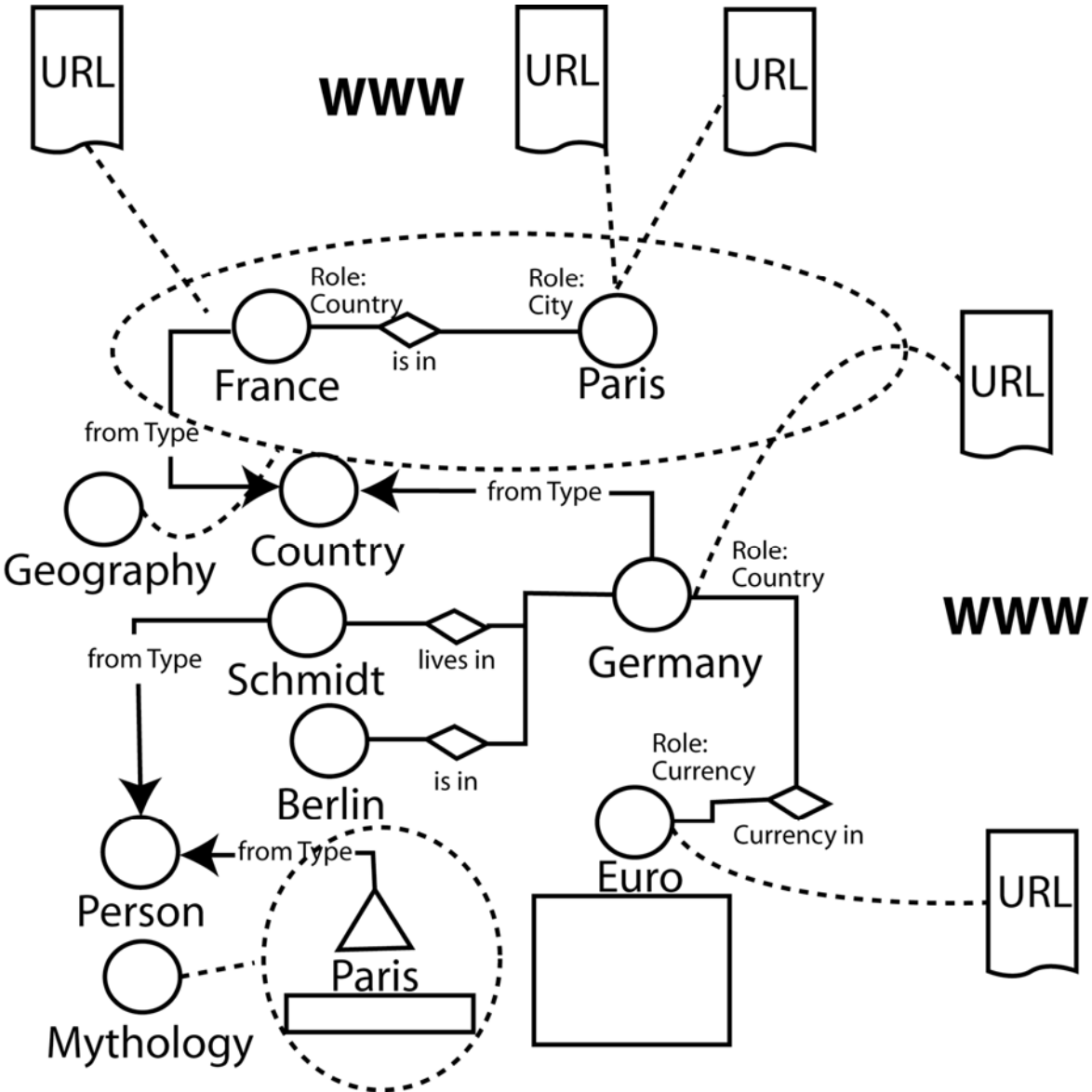
Entities of TopicMaps

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

- Topic
  - in hierarchy
  - Topic type
  - Public Subject Descriptor with identity attribute
  - Scope
- Occurrences: Links to external resources
- Associations: Relations between topics
- Facets: Name-value pairs attributed to Topics or Associations

TopicMaps - Example

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de



Ontologies

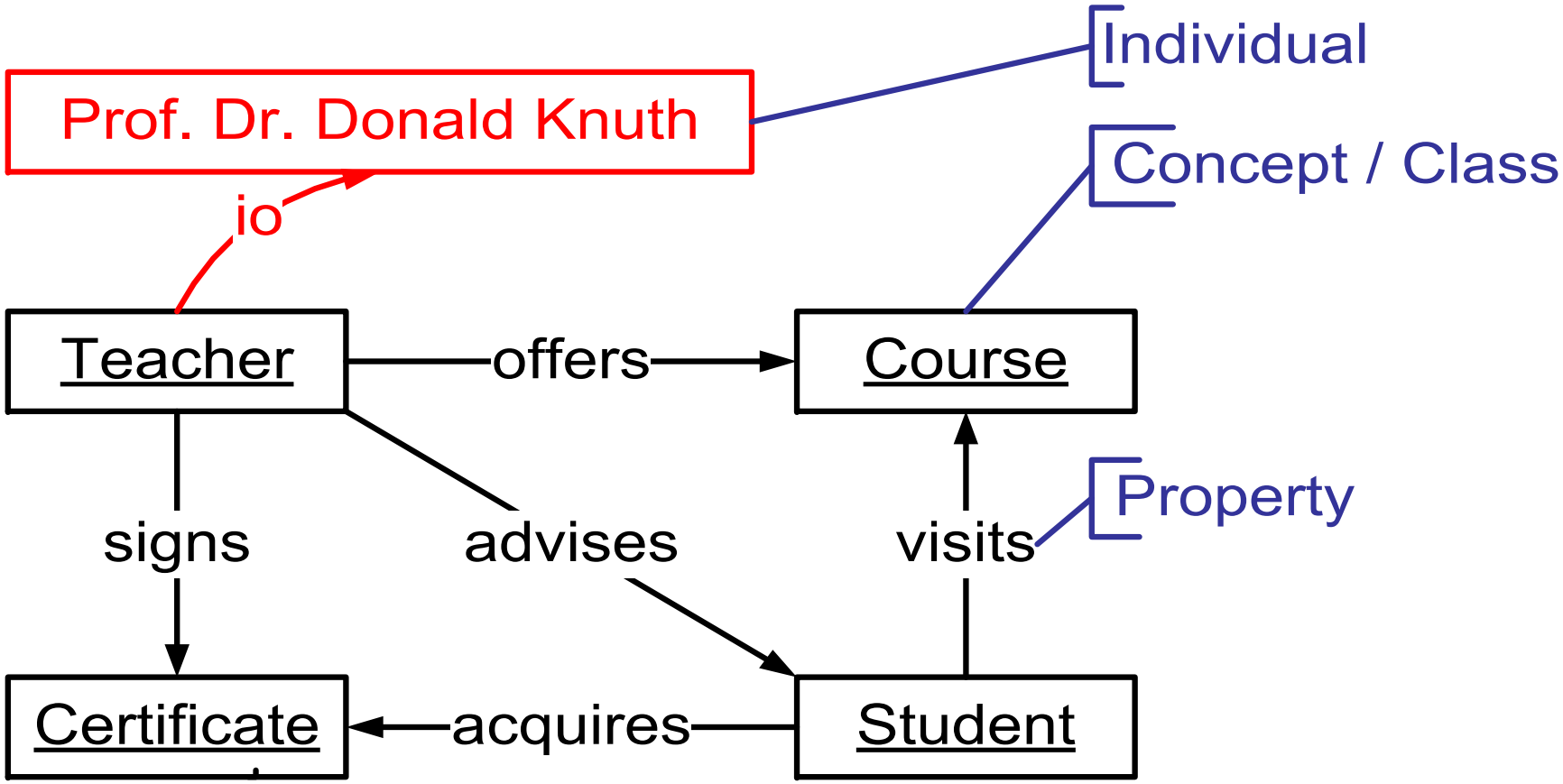
Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

- A formal, explicit specification of a shared conceptualization [Gruber93]
- Define formal semantics for information
- Define real-world semantics
- Pushed in artificial intelligence for knowledge sharing and re-use
- Description of knowledge domains:
  - Standardized terms (Classes, Axioms, etc.)
  - Relations between concepts
  - Inference rules

[Gruber93] http://ksl-web.stanford.edu/KSL_Abstracts/KSL-92-71.html

Ontology Example

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de



OWL

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

- Web Ontology Language:
  - W3C standard
  - Semantic markup language for publishing and sharing ontologies on the web
  - Successor of DAML+OIL
- Goals:
  - Mapping of Relations between vocables
  - Machine processable description of coherences
- Realisation: Extension of RDF

OWL Overview

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

- OWL Lite:
  - Simple expression of term hierarchies
  - Cardinality 0 or 1
- OWL DL (description logics):
  - Maximal expressivness while finitely computable
  - Some restrictions on nesting
- OWL Full:
  - Full expressiveness
  - No guarantied computablility

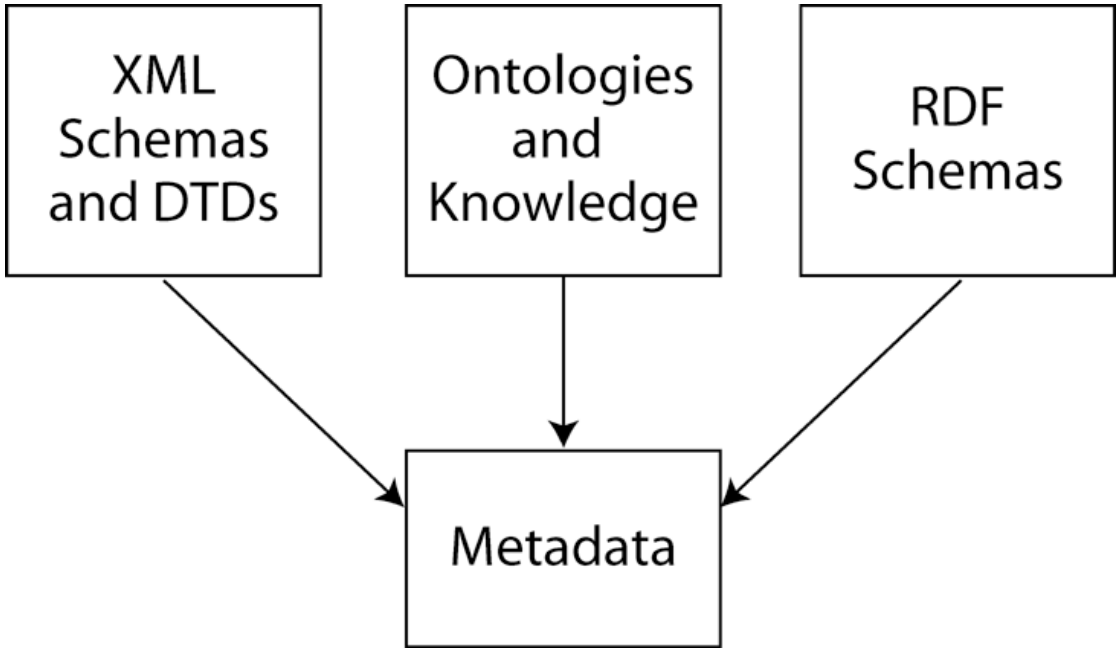
OWL Example

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

```
<owl:Class rdf:ID=''`Snake''>
   <rdfs:subClassOf rdf:resource=''#Animal''/>
</owl:Class>
<owl:Class rdf:ID=''Hamster''>
  <rdfs:subClassOf rdf:resource=''#Animal''/>
  <rdfs:subClassOf>
    <owl:Restriction>
     <owl:onProperty rdf:resource=''#hasParent''/>
     <owl:allValuesFrom rdf:resource=''#Hamster''/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <owl:disjointWith rdf:resource=''#Snake''/>
</owl:Class>
```

Redefined Notion of Metadata

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de



In the Semantic Web all available descriptors are needed as digestive Metadata!

Evaluating Resources

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

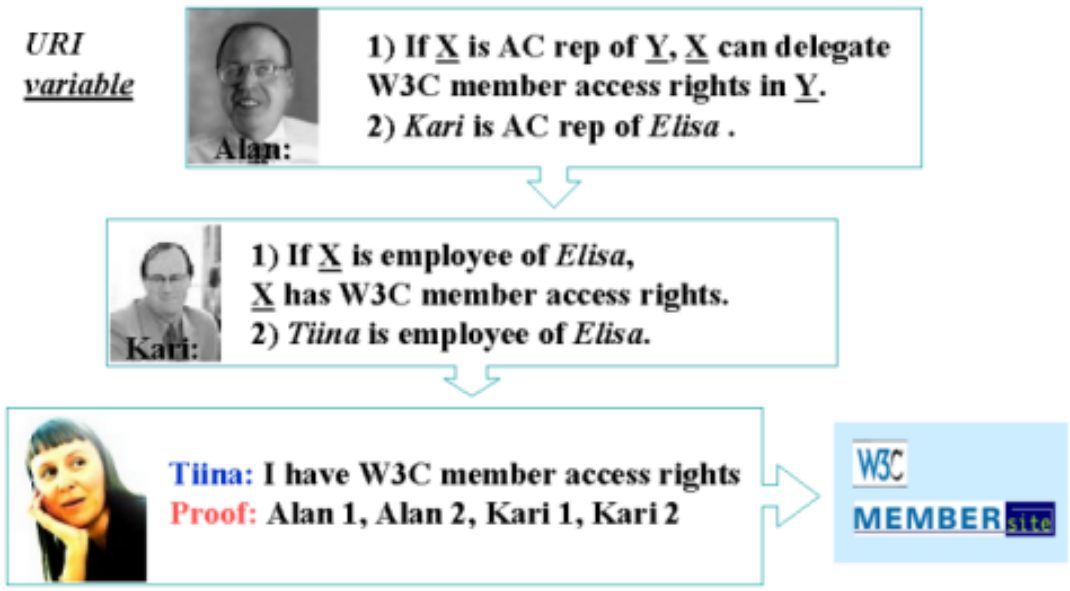
To judge on received information we need to evaluate:

- Hard states:
  - Authenticity of the source/author
  - Integrity of data
- Soft states:
  - Validity and reliability
  - Relevance
  - Context
  - Trustworthiness

W3C: Web of Trust

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

- Recipient determines group of trustees
- Trust can be inherited linearly according to rules
- Needs some certification (PKI, fingerprints …)
- Derived from the CA approach



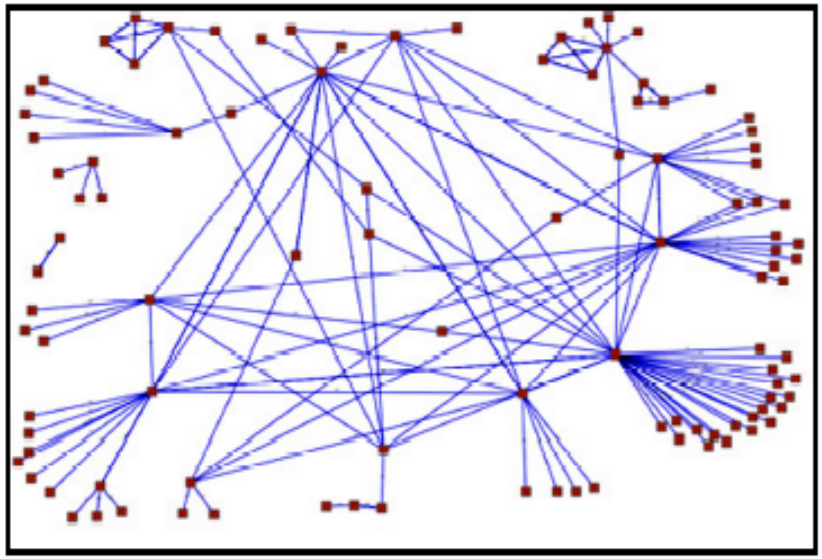
Poor Man's Logic: This approach is hierarchical - chains need unconditional trustworthiness at their roots!

Alternate Approach: Network Analysis

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

Explicit:

- Evaluate statements on your issue:
  "Dwayne you can trust"
  "Kilgore pays promptly"



Implicit:

- Evaluate statements and relations at hand
- Draw conclusions:
  "Donald Knuth is Professor at Stanford, thus I believe him."
  "Tim Berners-Lee is mentioned many times and in 'Network Hubs', he thus must be famous."

The Problem of Context

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

There are two contexts to consider

Context of creation:

- Donald Knuth writes on *Surreal Numbers* and *Diamond Signs*
- R. Gernhardt links the words “My Favourite” to M. Reich-Ranickis Book page from a paragraph titled “Most Awkward Publications”

Context of reception:

- “I know the works of Knuth, but am looking for young talents”
- “1 billion of Chinese think that something is good, but my Grandma does not”

Problem: Identify contexts and judge on their compatibility/agreement.

The Problem of Time

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

Reputation of a resource (person, institution, agent, …) is a function of time …

Example: Konrad Zuse, the well reputed pioneer of computer systems, published papers ‘of lesser renown’ on fundamental physics in its later age.

In general: The reputation of a resource is an expectation about its current behaviour based on information about or observations of its past behaviour.

The Problem of Induced Biases

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

Implicit:

- Structural inheritance: URI of D. Knuth's homepage could be of identical structural formation as some technical staff (is not)
- General problem: How to account for the deep Web

Explicit:

- Trust inflation: People/institutions granting plentiful amounts of reputations
- Destructive groups: Groups injecting 'consistent falseness' on large scale
- Large players: Players owning many Web sites may enforce self-exaltation
- Software vendors/pirates: Leading software vendors (or software pirates) may (self-)reinforce by 'default settings'

Résumé on Resource Evaluation, Reputation & Trust

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

- Not simple at all
- Would like a global PKI
- Suffers from conceptual unclearness in basic Web semantics: Contexts, Links, …
- Suffers from the certainty about a persistent structural chaos in the Web
- Some promising heuristics
- Active research area

References

Thomas Schmidt
schmidt@informatik.
haw-hamburg.de

- Semantic Web @ W3C - http://www.w3.org/2001/sw/
- Semantic Web – http://www.semanticweb.org/
- Daconta, Obrst, Smith: *The Semantic Web*, Wiley 2003.
- B. Thuraisingham: *XML Databases and the Semantic Web*, CRC Press, 2002.
- Ubbo Visser et. al: Web Development, WWW Tutorial May 2004
- R. Widhalm, Th. Mück: *Topic Maps*, Springer 2002.
- D. Fensel: *Ontologies*, Springer 2004.
- RDF Primer - http://www.w3.org/TR/rdf-primer/
- RDF Model & Abstract Syntax - http://www.w3.org/TR/rdf-concepts/
- Dublin Core – http://dublincore.org
- DDC - http://www.oclc.org/dewey/default.htm
  http://www.gutenberg.net/dirs/1/2/5/1/12513/12513-h/12513-h.htm