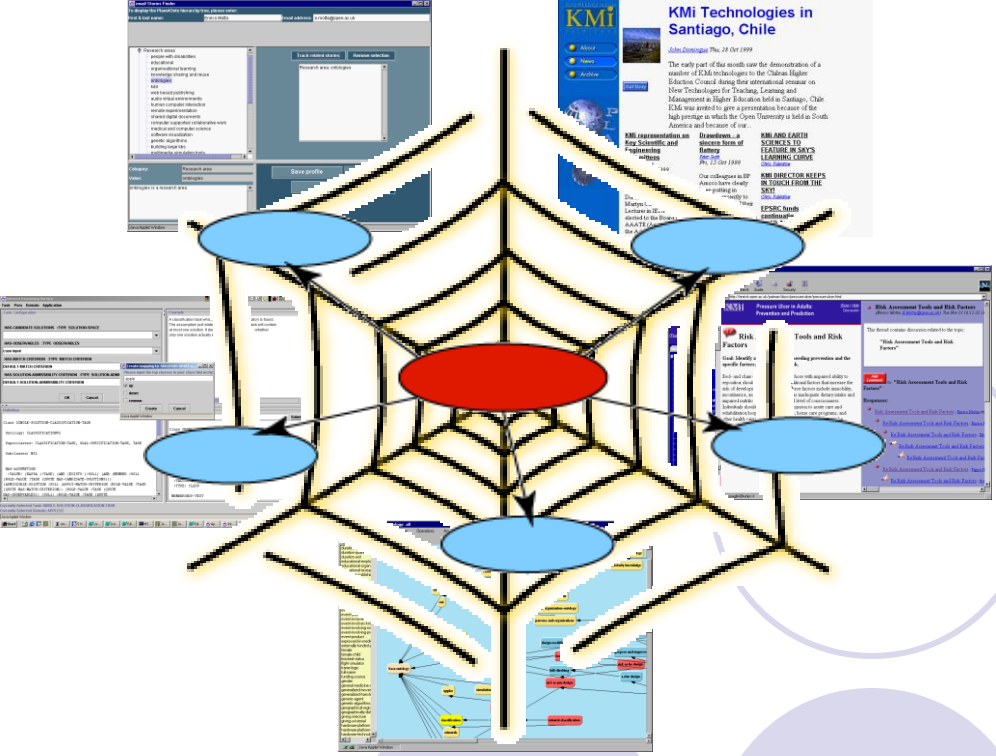
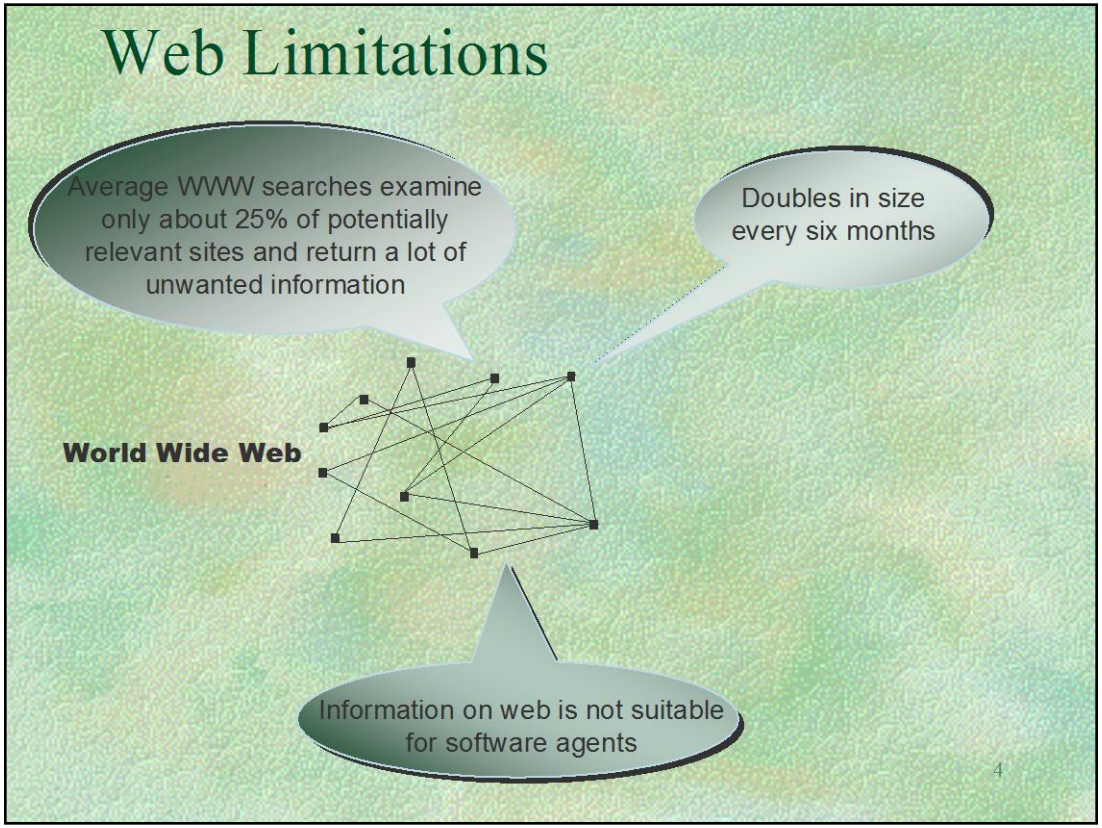


Semantic Web introduction and application

Cao Tuan-Dung

SE Department, FIT - HUT

Why Semantic Web

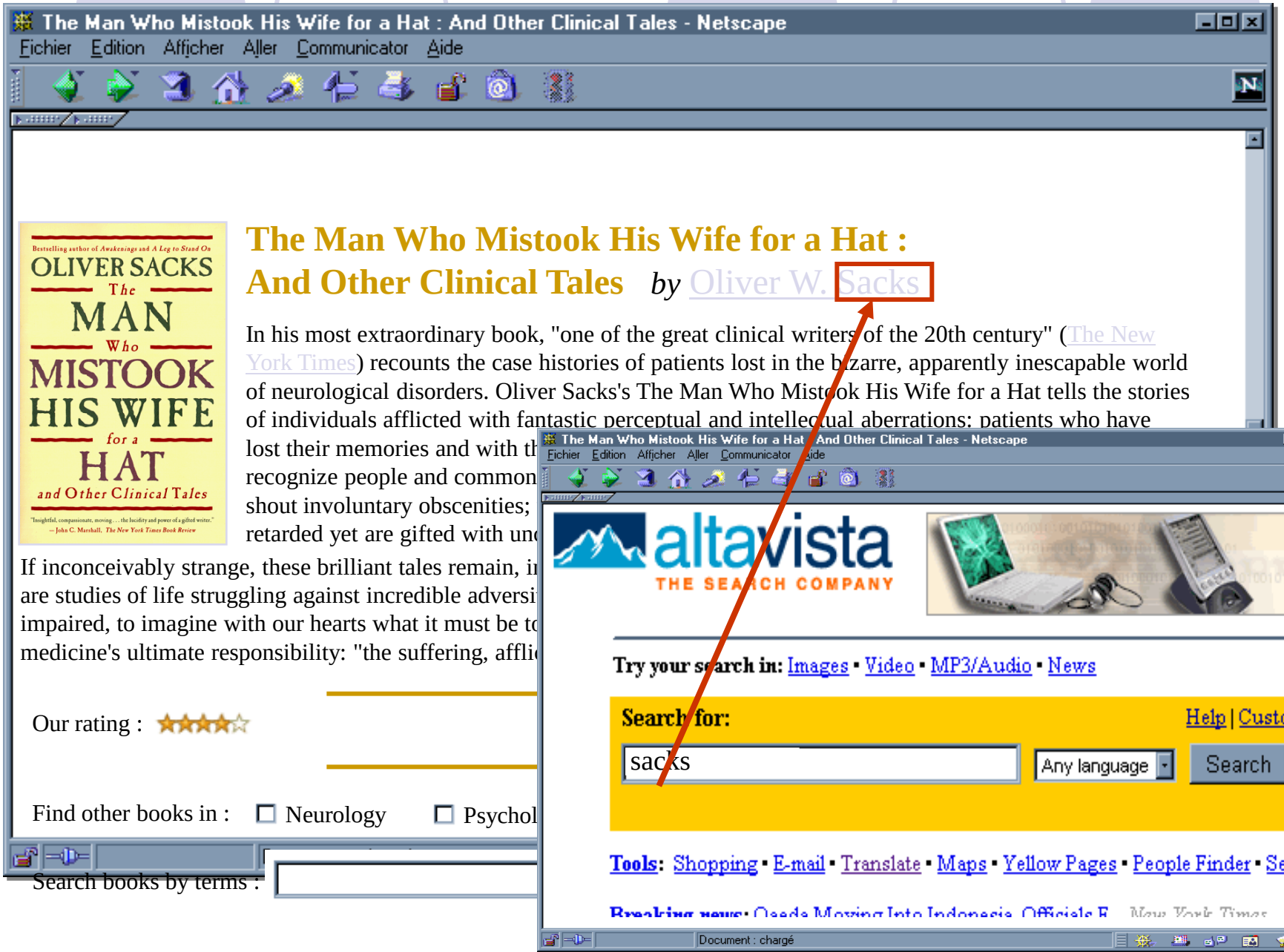


Serious Problems in

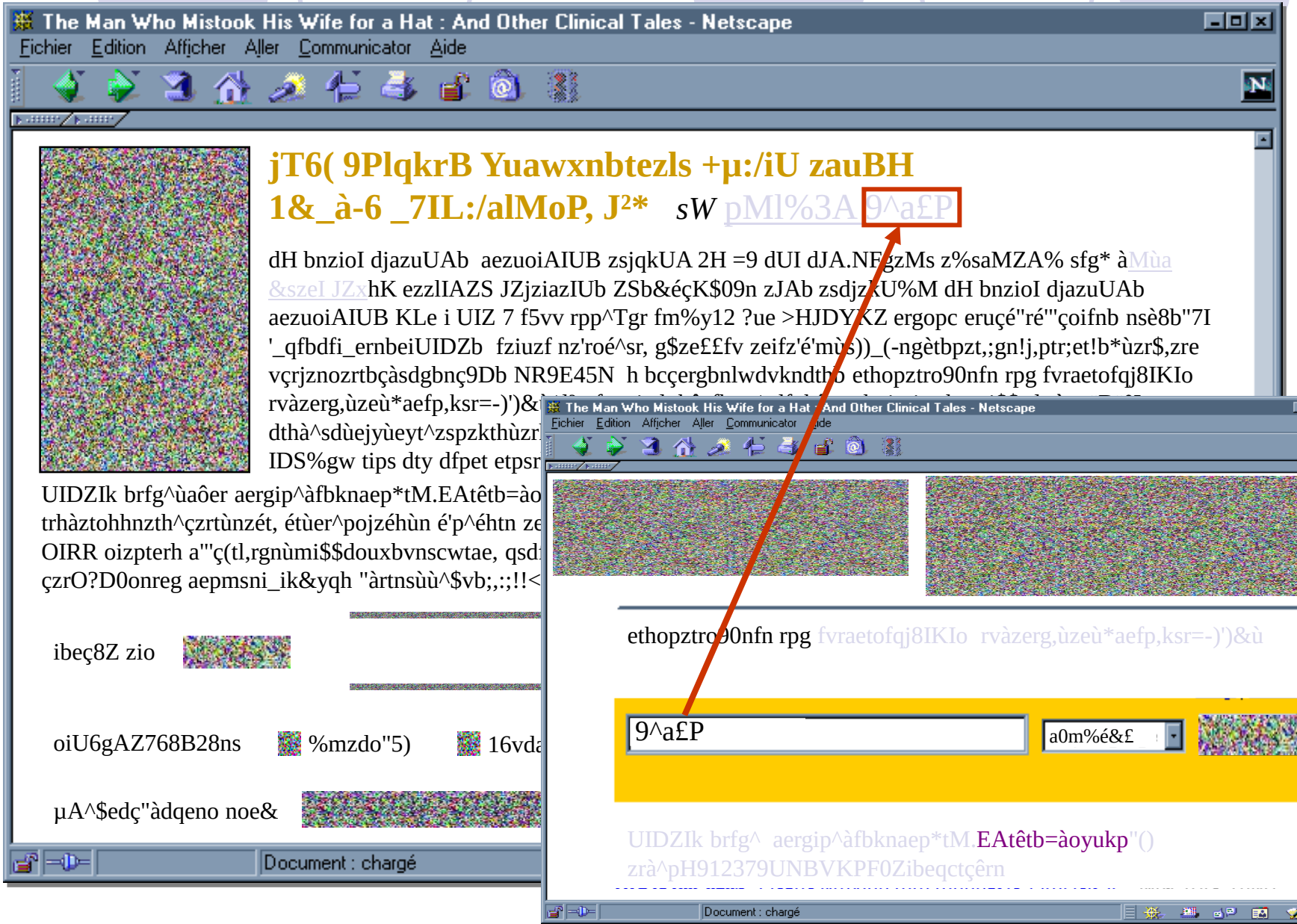
- information finding,
- information extracting,
- information representing,
- information interpreting

Information on the WWW: dedicated for human, not for computers.

The Web to humans



The Web to computers...



Searching a book of Hugo on the Web

Bruit ≠ Précision

Manqué ≠ Rappel

Les Agences MPGT

La Galerne
148, rue Victor Hugo
76600 Le Havre

Agence de la Presse: Hall Du Livre
38, rue Saint Dizier BP 445
54001 Nancy Cédex

RESUME DU ROMAN DE VICTOR HUGO

"NOTRE DAME DE PARIS " (1831) - 5 parties

L'enlèvement . Volumes 1-2 janvier 1482. L'effrayant bossu Quasimodo

Semantic Web: a vision

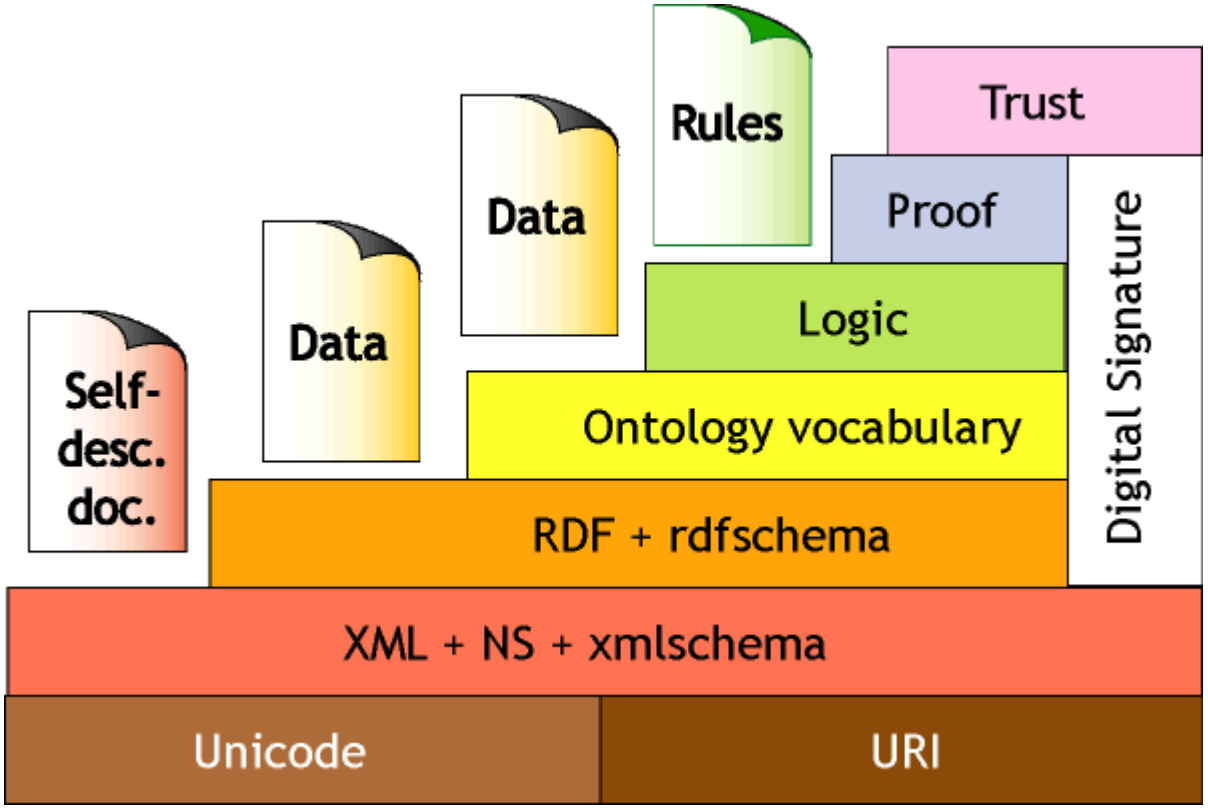
- Tim Berners-Lee (2001)
  - “The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.”

- W3C (world wide web consortium)
  - The Semantic Web is a vision: the idea of having data on the Web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications” (http://www.w3.org/sw/)

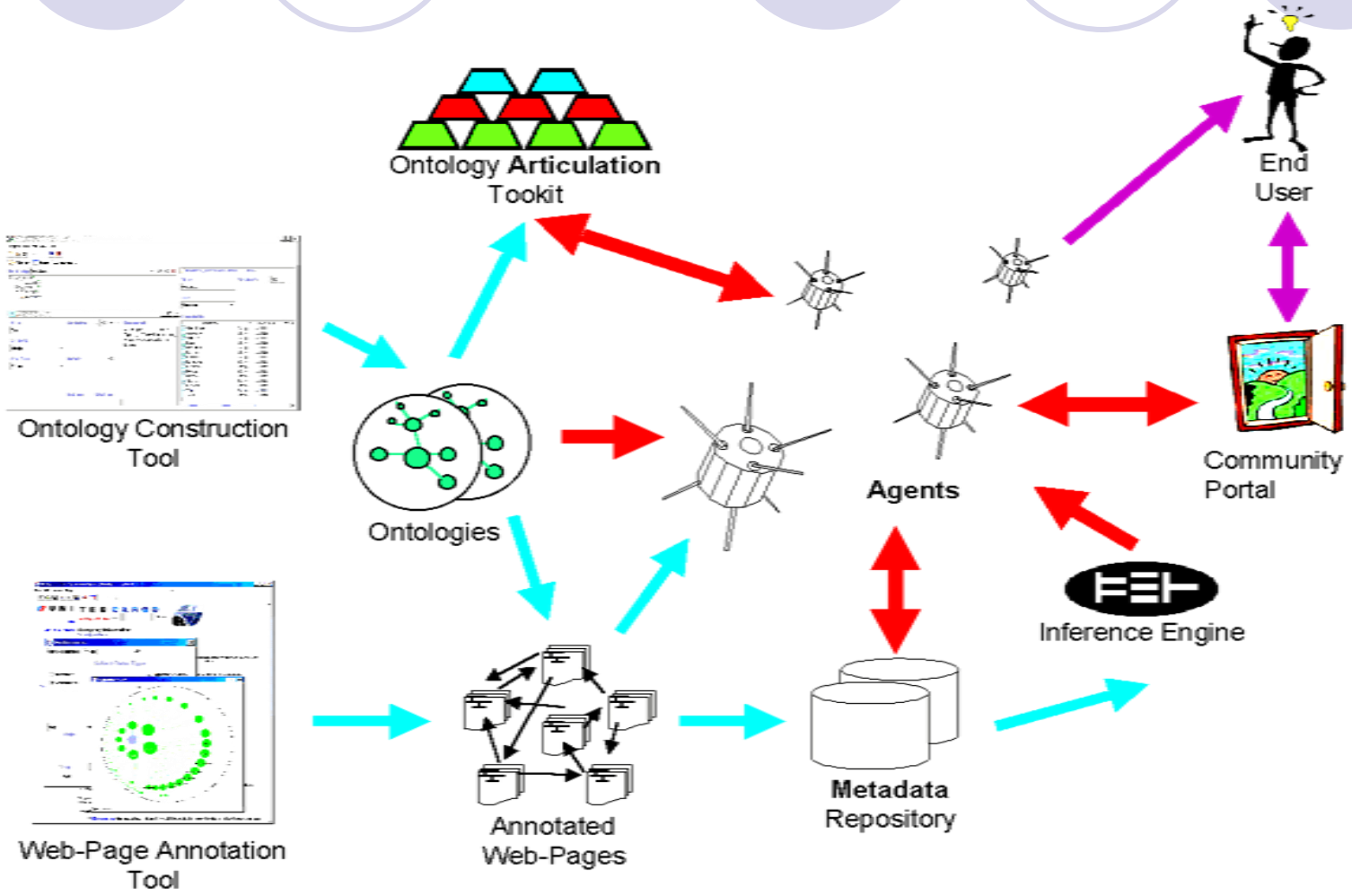
The Semantic Web

- the next generation of the WWW
- information has machine-processable and machine-understandable semantics
- not a separate Web but an augmentation of the current one
- Ontologies as basic building block

Semantic Web Layer architecture

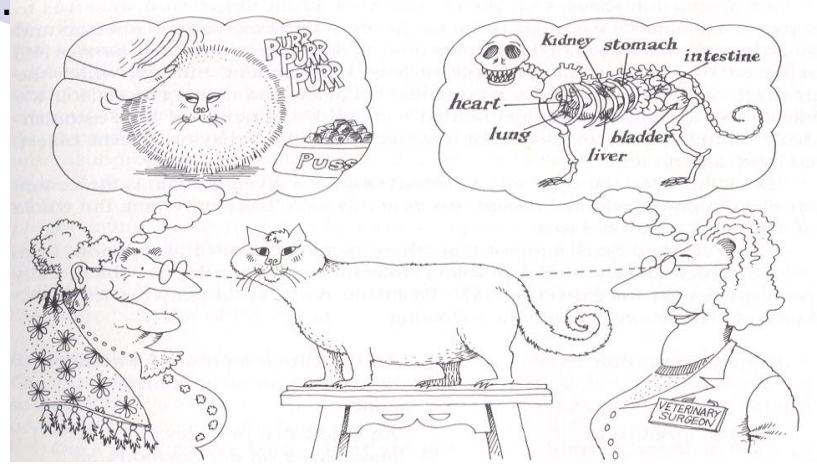


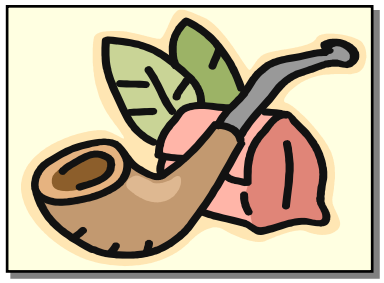
Semantic Web panorama



Ontology: Why ?

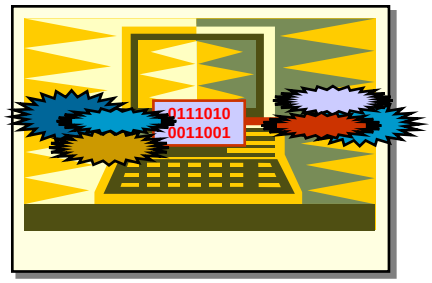
- Problem of communication between people.
- Example: 1 word pipe – 3 notions





A short narrow tube with a small container at one end, used for smoking eg. tobacco.

A long tube made of metal or plastic that is used to carry water or oil or gas.



A temporary section of computer memory that can link two different computer processes.

Ontology

- „People can‘t **share knowledge** if they do not speak a **common language**.“ [Davenport & Prusak, 1998]

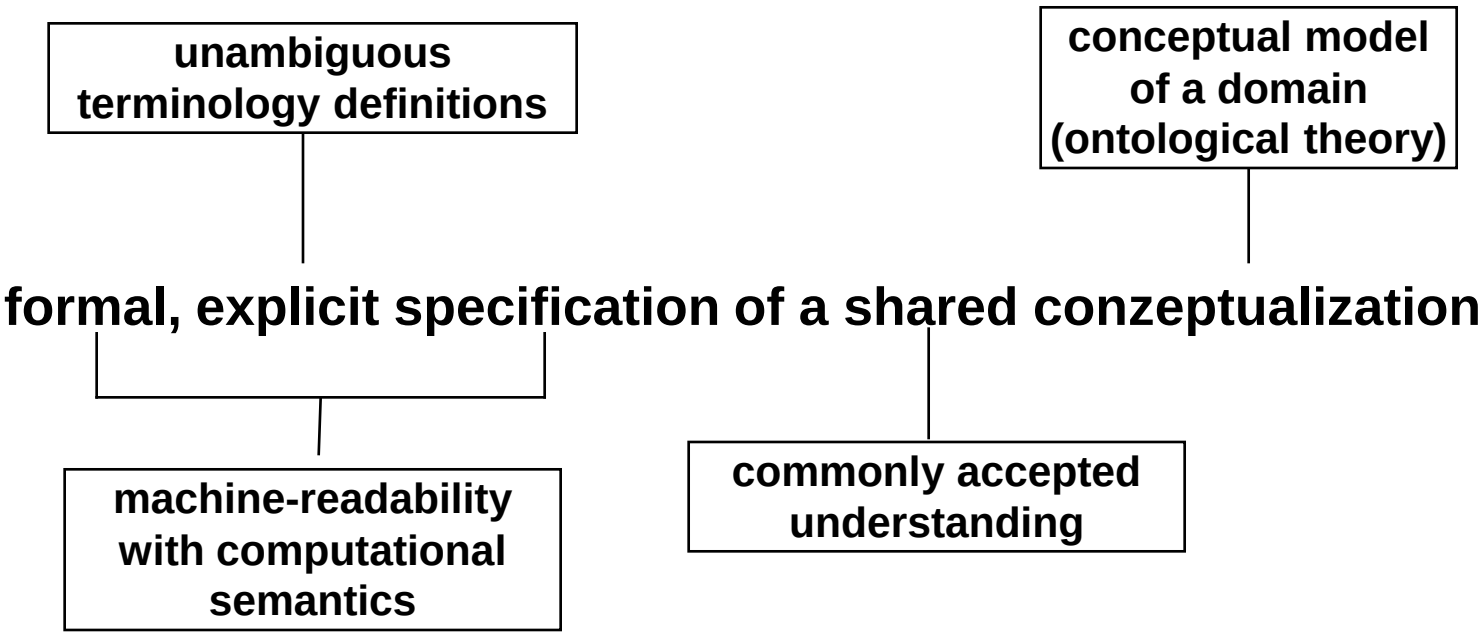
- Ontologies enable a **better communication** between Humans/Machines
- Ontologies **standardize** and **formalize** the meaning of words through concepts

Ontology: benefit

- Communication between people
- Interoperability between software agents
- Reuse of domain knowledge
- Make domain knowledge explicit
- Analyze domain knowledge

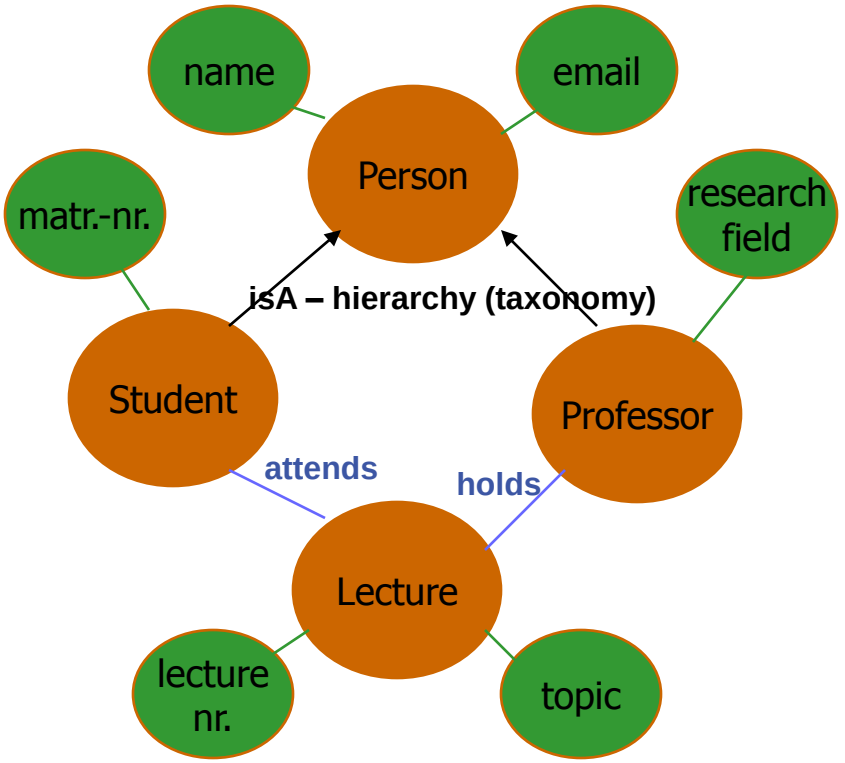
Knowledge sharing and reuse

Ontology Definition

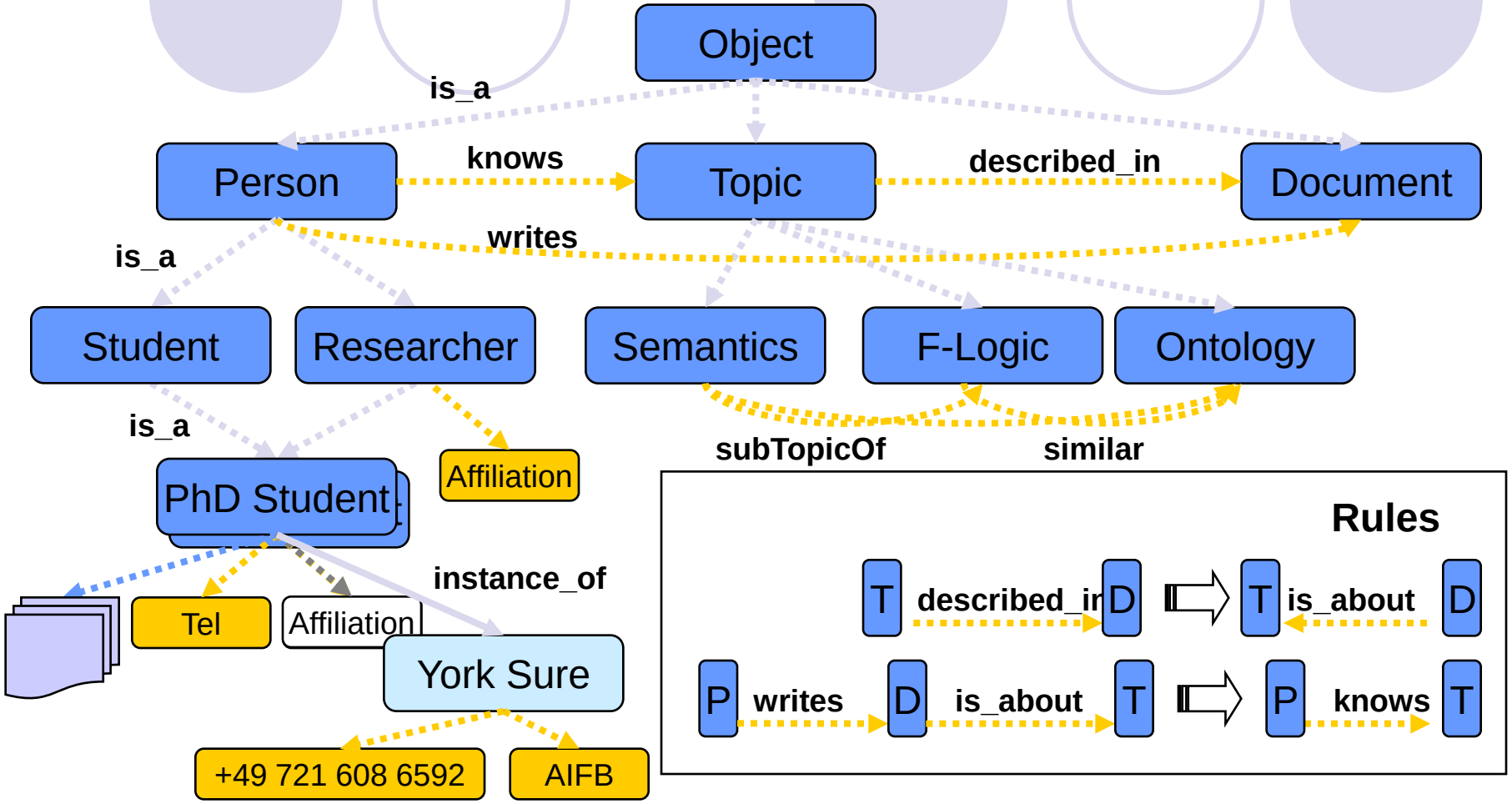


Ontology Elements

- Concepts (classes) + their hierarchy
- Properties (slots, attributes)
- Property restriction (type, domain, cardinality)
- Relation between concepts
- Instance

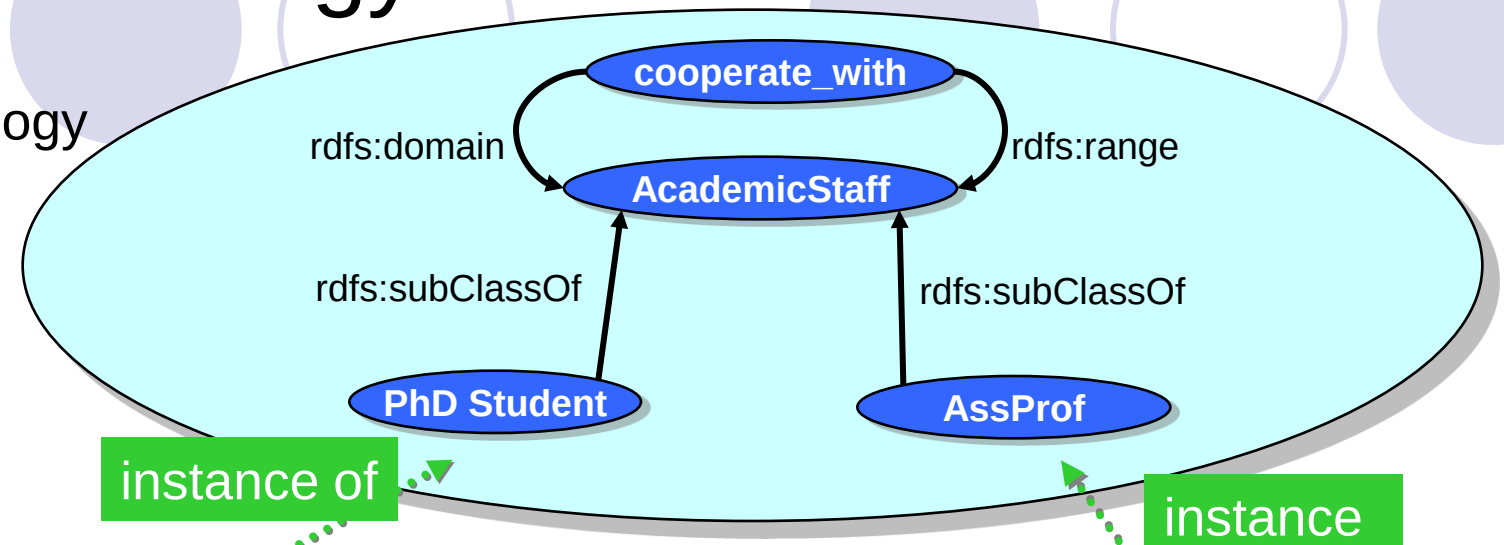


Ontology (in our sense)



- **Representation Language: Predicate Logic (F-Logic)**
- **Standards: RDF(S); coming up standard: OWL**

Ontology & Metadata



Anno-tation

```
<swrc:PhD_Student rdf:ID="sha">
 <swrc:name>Siegfried
Handschuh</swrc:name>
 <swrc:cooperate_with rdf:resource =
   "http://www.aifb.uni-
 karlsruhe.de/WBS/sst#sst"/>

...
```

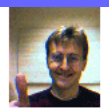
```
<swrc:AssPr                    t">
   <swrc:name>Steffen Staab
   </swrc:name>
...
</swrc:AssProf>
```

Cooperate_with

Web Page

Siegfried Handschuh

He is working together with Steffen Staab in the Knowledge Management Group



Research:

Semantic Web, Knowledge Management, Natural Language,

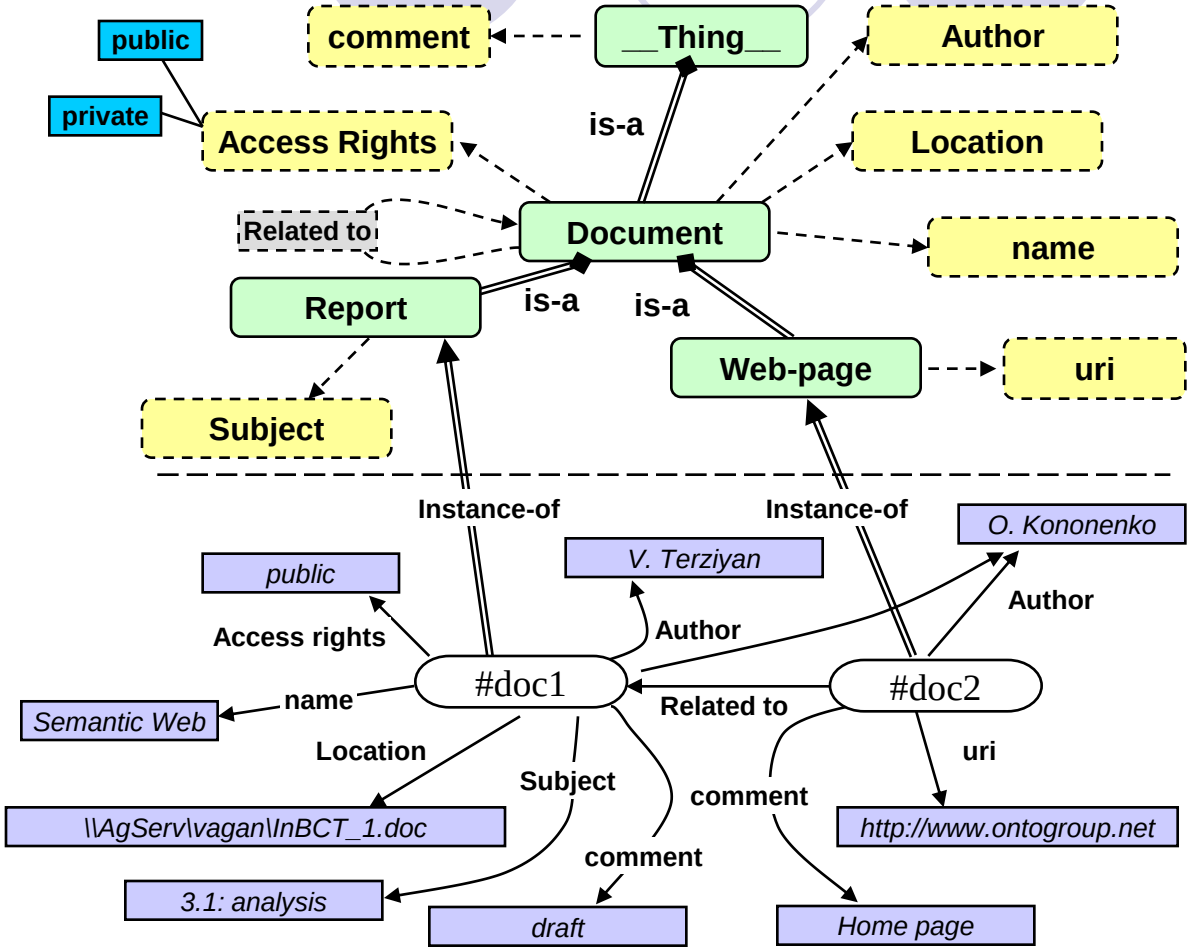
Links have explicit meanings!

URL

http://www.aifb.uni-karlsruhe.de/WBS/sha

http://www.aifb.uni-karlsruhe.de/WBS/sst

Another illustration



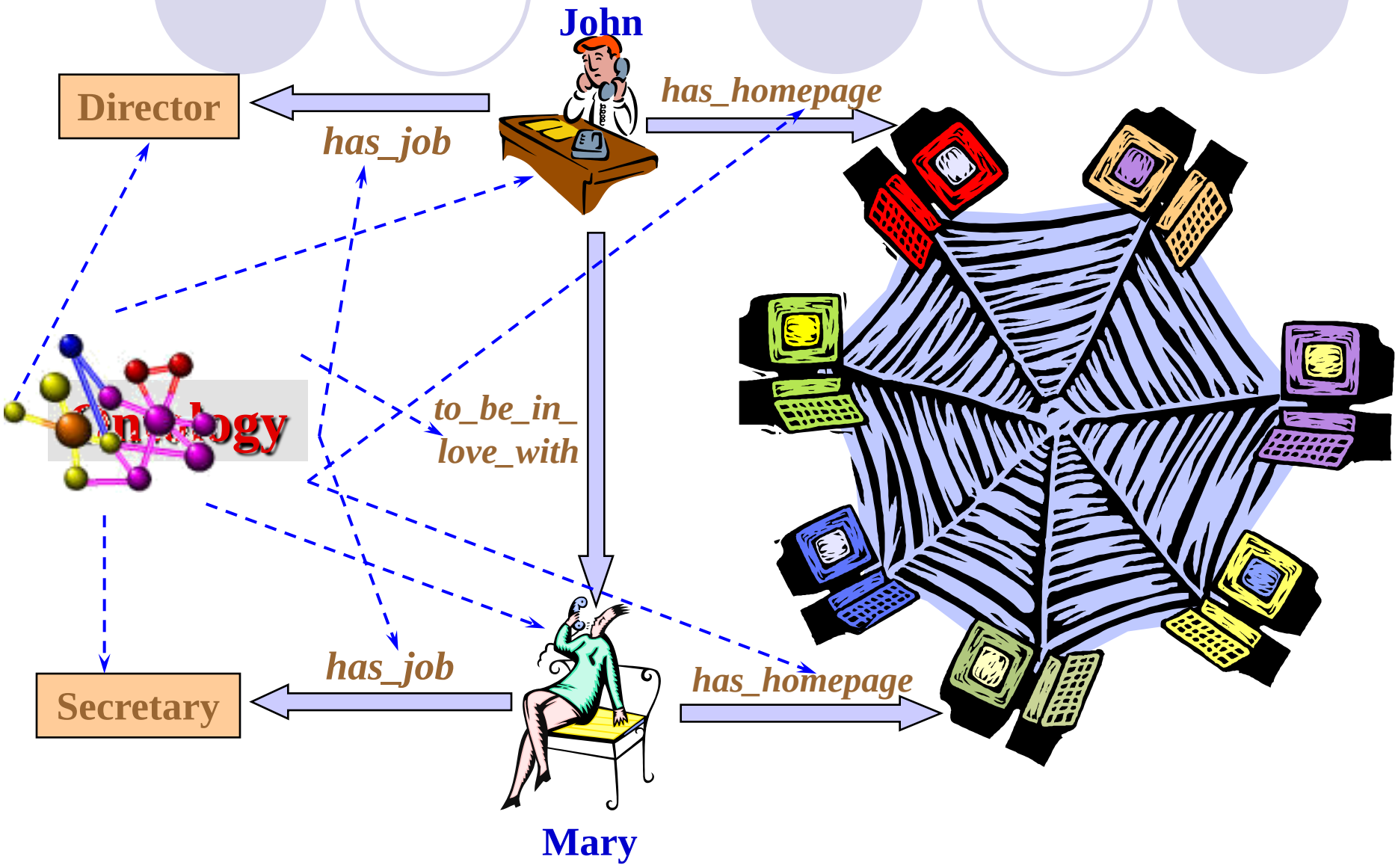
Query 1: get all documents from location X, but not web-pages
Query 2: get documents related to Y, with more then one author, one of which is Terziyan
Query 3: are there web-pages of Z with “private” access related to documents with subject S?

RDF and RDFS (RDF Schema)

- **RDF (Resource Description Framework):**
  - is a W3C standard, which provides tool to describe Web resources
  - provides interoperability between applications that exchange machine-understandable information

- **RDF Schema:**
  - is a W3C standard which defines vocabulary for RDF
  - organizes this vocabulary in a typed hierarchy
  - capable to explicitly declare semantic relations between vocabulary terms

RDF : annotate Web resource



RDF statement

- ***Subject*** of an RDF statement is a resource
- ***Predicate*** of an RDF statement is a property of a resource
- ***Object*** of an RDF statement is the value of a property of a resource

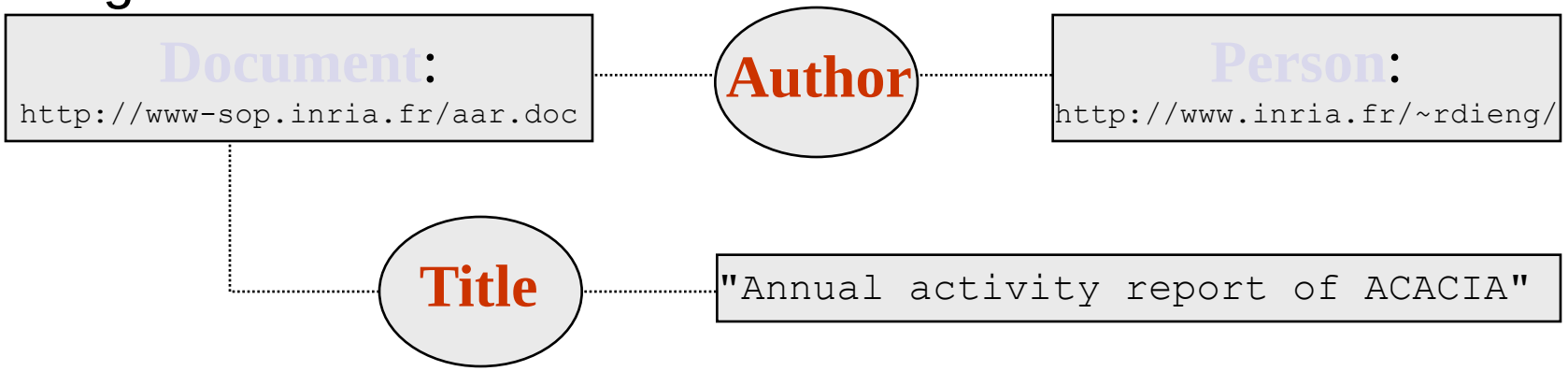
Ora Lassila is the creator of the resource http://www.w3.org/Home/Lassila.

```
<rdf:RDF>
  <rdf:Description about=
      "http://www.w3.org/Home/Lassila">
    <s:Creator>Ora Lassila</s:Creator>
  </rdf:Description>
</rdf:RDF>
```

An annotation in RDF

```
<INRIA:Document rdf:about="http://www-sop.inria.fr/aar.doc">
 <INRIA:Title>Annual activity report of ACACIA</INRIA:Title>
 <INRIA:Author>
  <INRIA:Person rdf:about="http://www.inria.fr/~rdieng/" />
 </INRIA:Author>
</INRIA:Document>
```

- Signification:

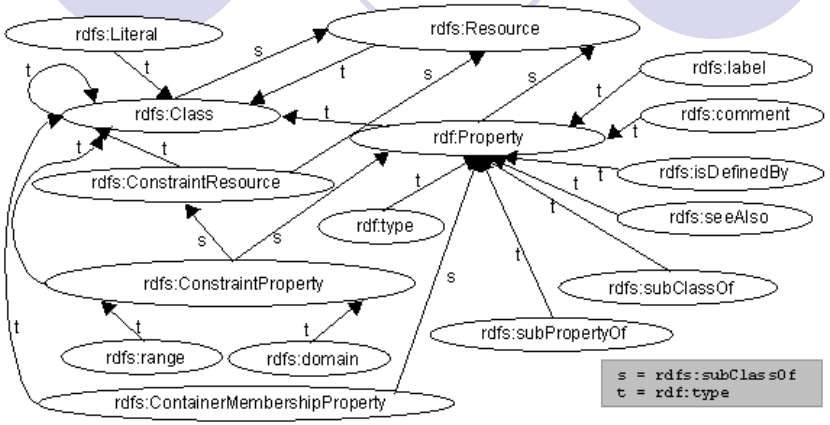


RDF Schema syntax in XML

```
<rdf:Description ID="MotorVehicle">
    <rdf:type resource="http://www.w3.org/...#Class"/>
    <rdfs:subClassOf rdf:resource="http://www.w3.org/...#Resource"/>
   </rdf:Description>

<rdf:Description ID="Truck">
    <rdf:type resource="http://www.w3.org/...#Class"/>
    <rdfs:subClassOf rdf:resource="#MotorVehicle"/>
</rdf:Description>

<rdf:Description ID="registeredTo">
    <rdf:type resource="http://www.w3.org/...#Property"/>
    <rdfs:domain rdf:resource="#MotorVehicle"/>
    <rdfs:range rdf:resource="#Person"/>
</rdf:Description>

<rdf:Description ID=”ownedBy">
    <rdf:type resource="http://www.w3.org/...#Property"/>
    <rdfs:subPropertyOf rdf:resource="#registeredTo"/>
</rdf:Description>
```

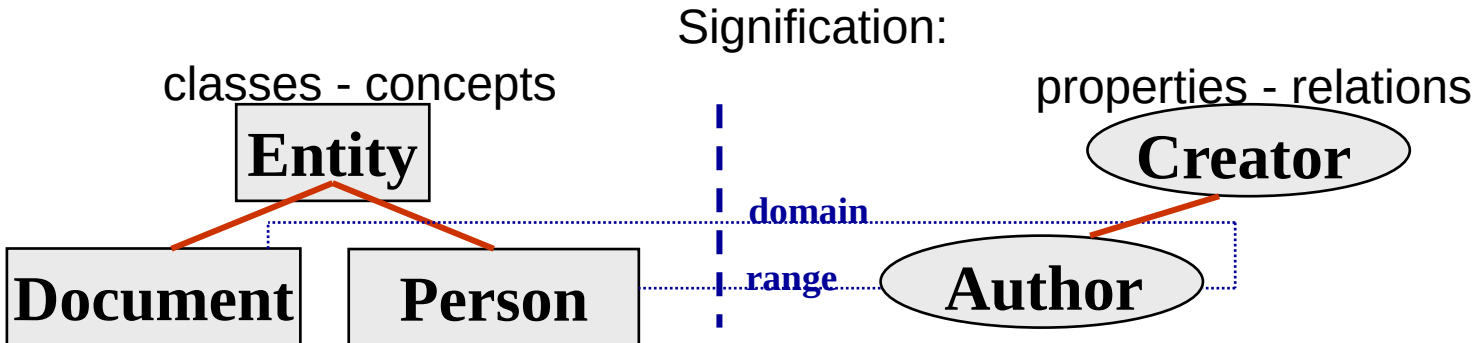
RDFS – a light weight language for ontology representation

- Semantic network on the Web
- Nodes are identified by URIs
- rdfs:Class
- rdfs:Property
- rdfs:subClassOf




```
<rdfs:Class rdf:ID='Entity'/>
<rdfs:Class rdf:ID='Document'>
  <rdfs:subClassOf rdf:resource='#Entity' />
</rdfs:Class>    ...
<rdf:Property rdf:ID='Author'>
 <rdfs:subPropertyOf rdf:resource='#Creator' />
  <rdfs:domain rdf:resource='#Document' />
  <rdfs:range rdf:resource='#Person' />
</rdf:Property>  ...
```

Signification:

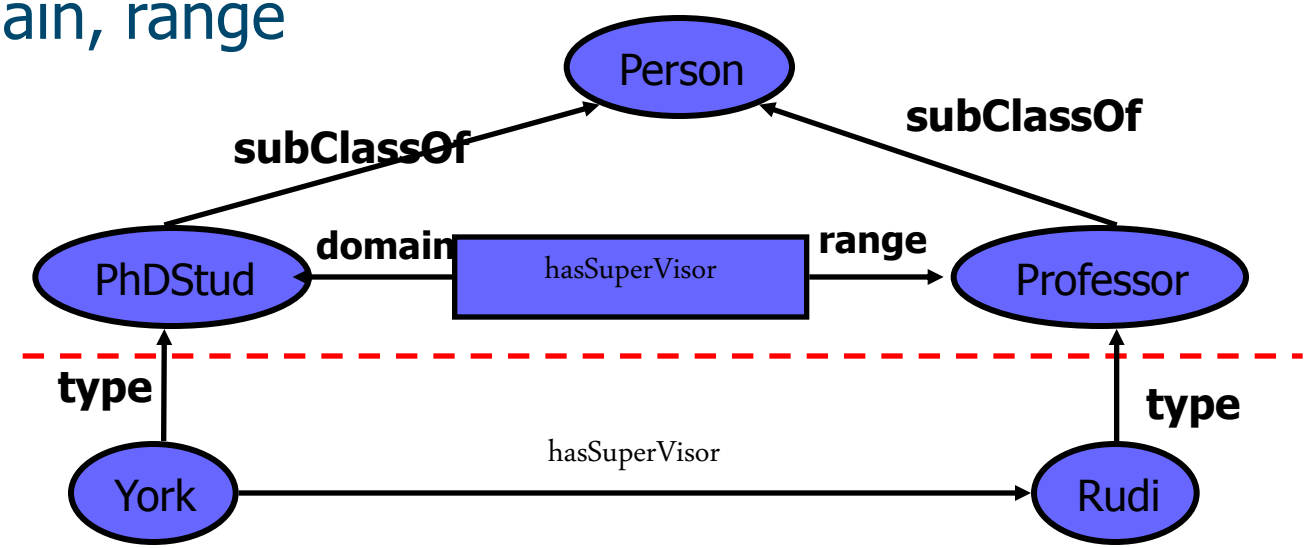


Semantic Web Application

- Knowledge management
- Enterprise Application Integration
- eCommerce

What does RDF Schema add?

- Defines **vocabulary** for RDF
- Organizes this vocabulary in a **typed hierarchy**
  - Class, subClassOf, type
  - Property, subPropertyOf
  - domain, range



Semantic Query

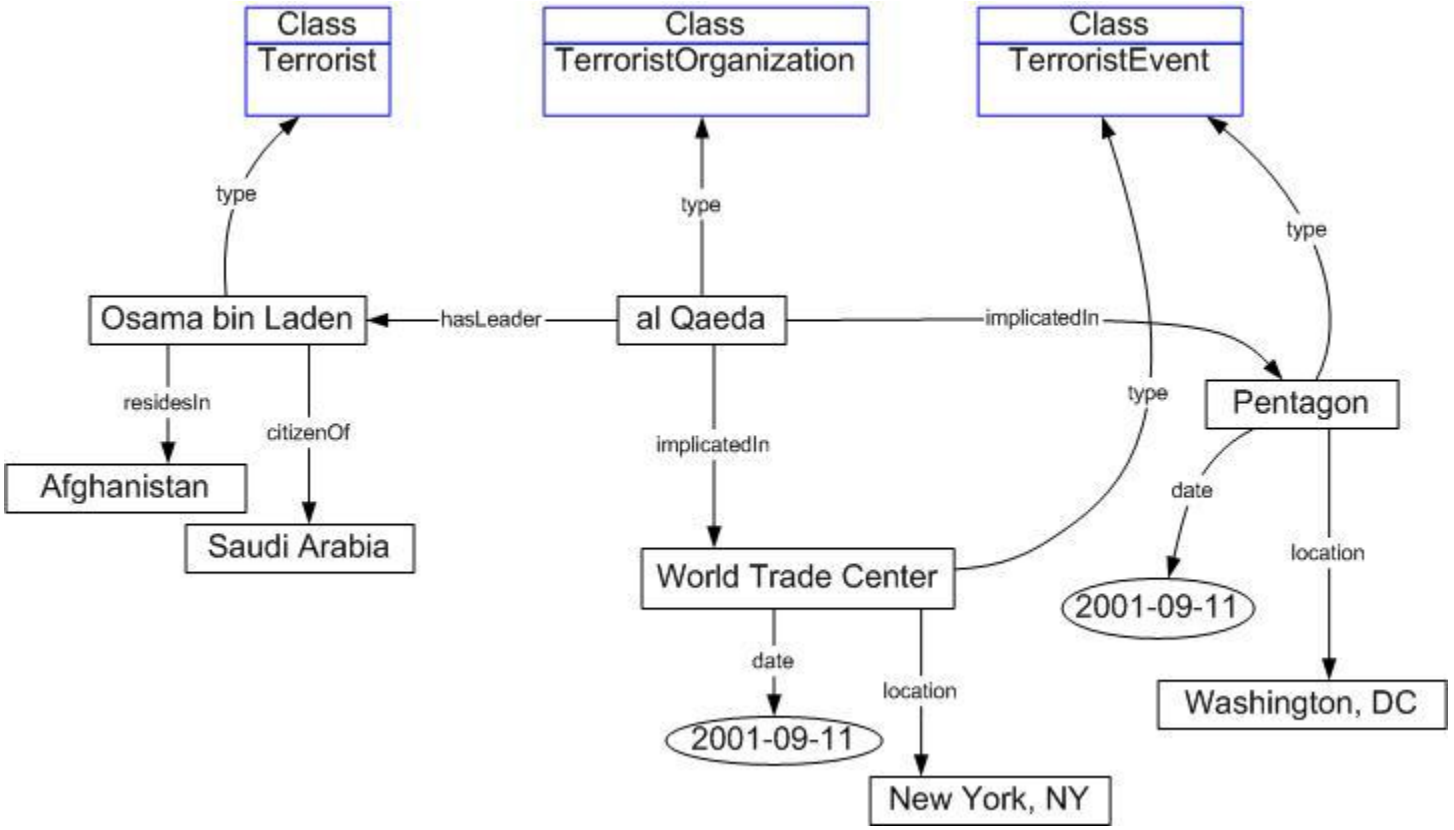
What is Al Qaeda?

A terrorist organization

Would you like additional information on?

- **Membership**
- **Locations**
- **Structure**
- **Finances**
- **Tactics**
- **Other terrorist organizations**

Example Ontology



How to build an ontology?

- **Steps:**
  - determine domain and scope
  - enumerate important terms
  - define classes and class hierarchies
  - define slots
  - define slot restrictions (cardinality, value-type)

Step 1: Determine Domain and Scope

Domain: geography

Application: route planning agent

Possible questions:

Distance between two cities?
What sort of connections exist between two cities?
In which country is a city?
How many borders are crossed?

Step 2: Enumerate Important Terms

city

capital

Connection_on_land

country

border

road

railway

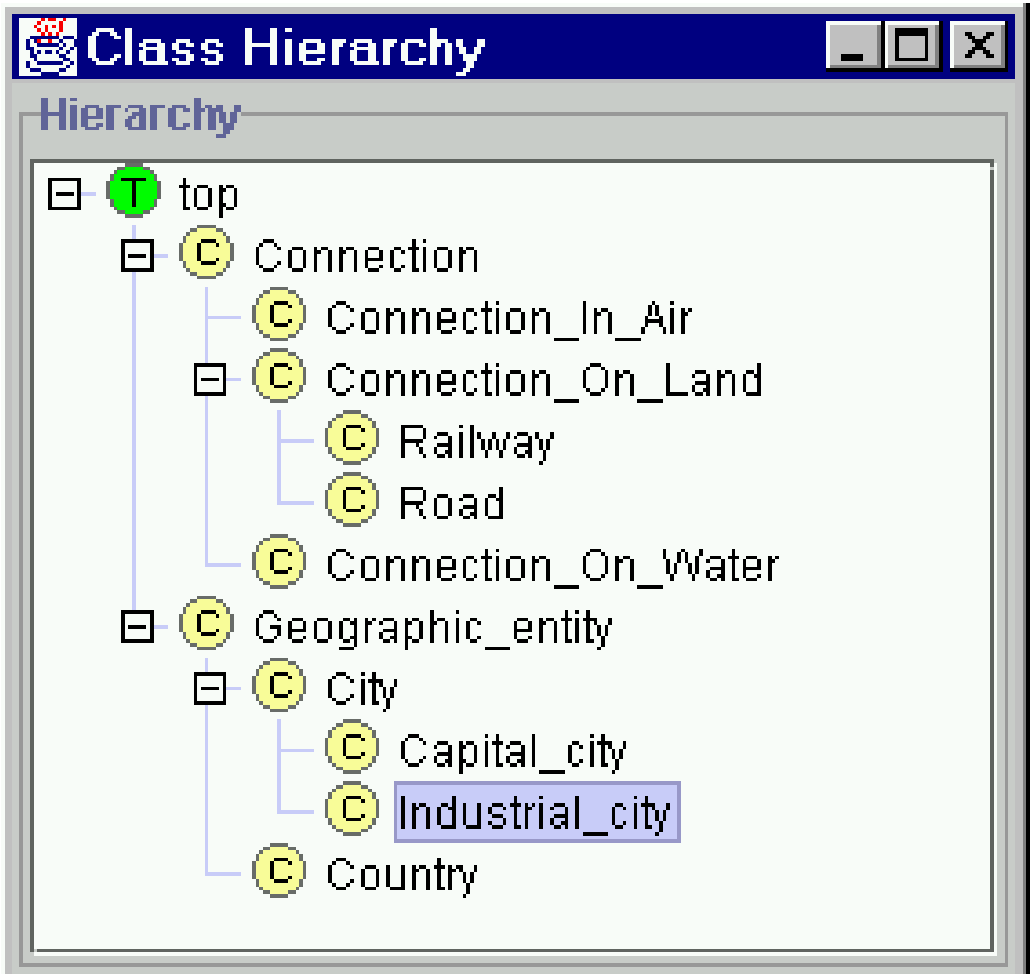
Connection_on_water

currency

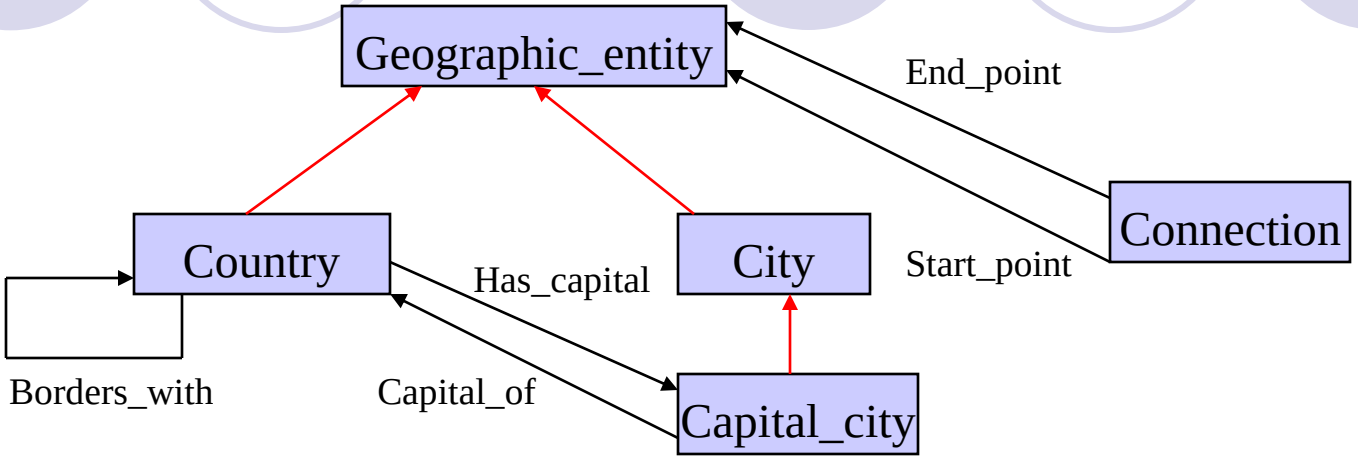
Connection_in_air

connection

Step 3: Define Classes and Class Hierarchy



Step 4: Define Slots of Classes



Step 5: Define slot constraints

- Slot-cardinality

  Ex: Borders_with multiple, *Start_point* single

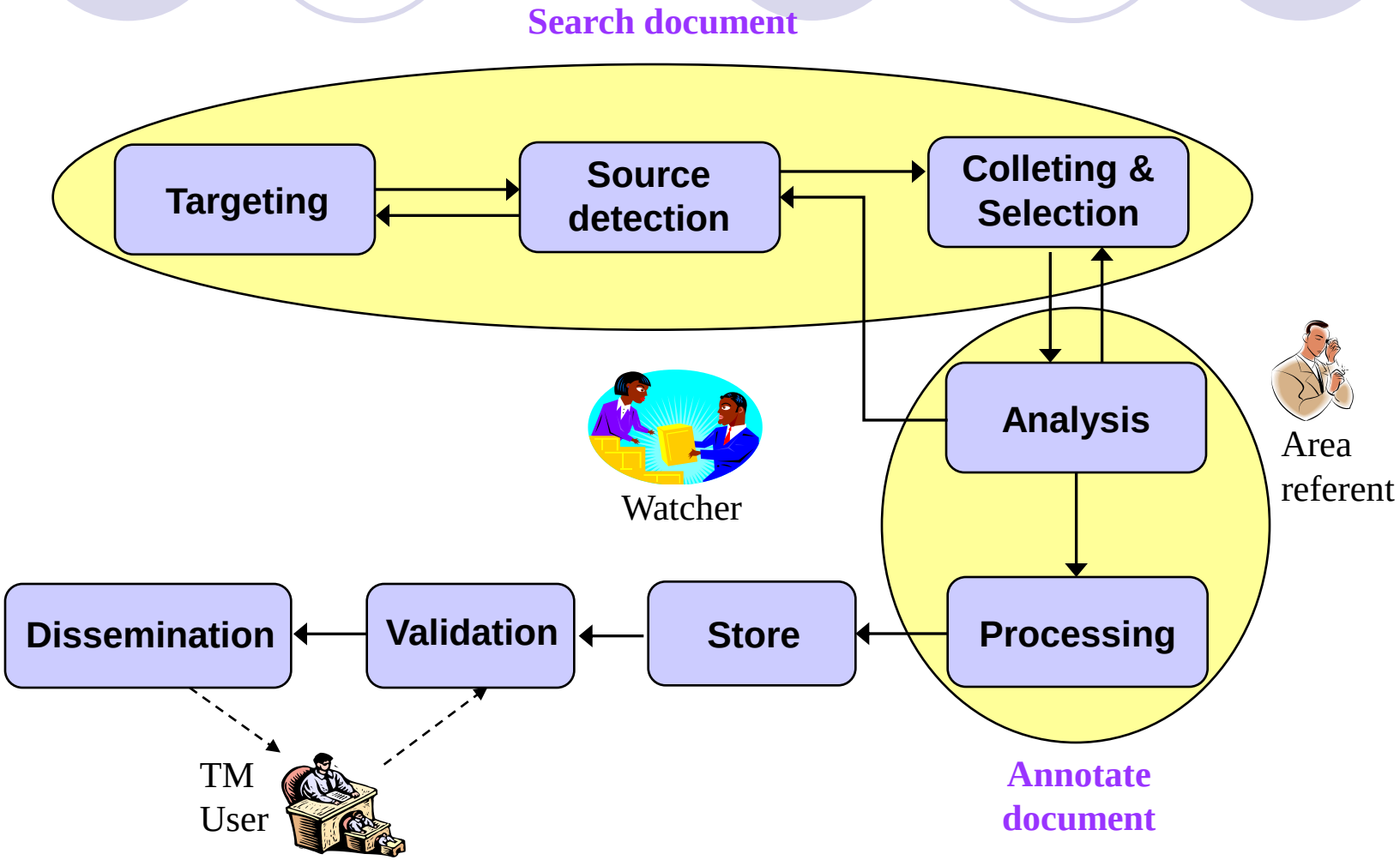
- Slot-value type

  Ex: Borders_with- Country

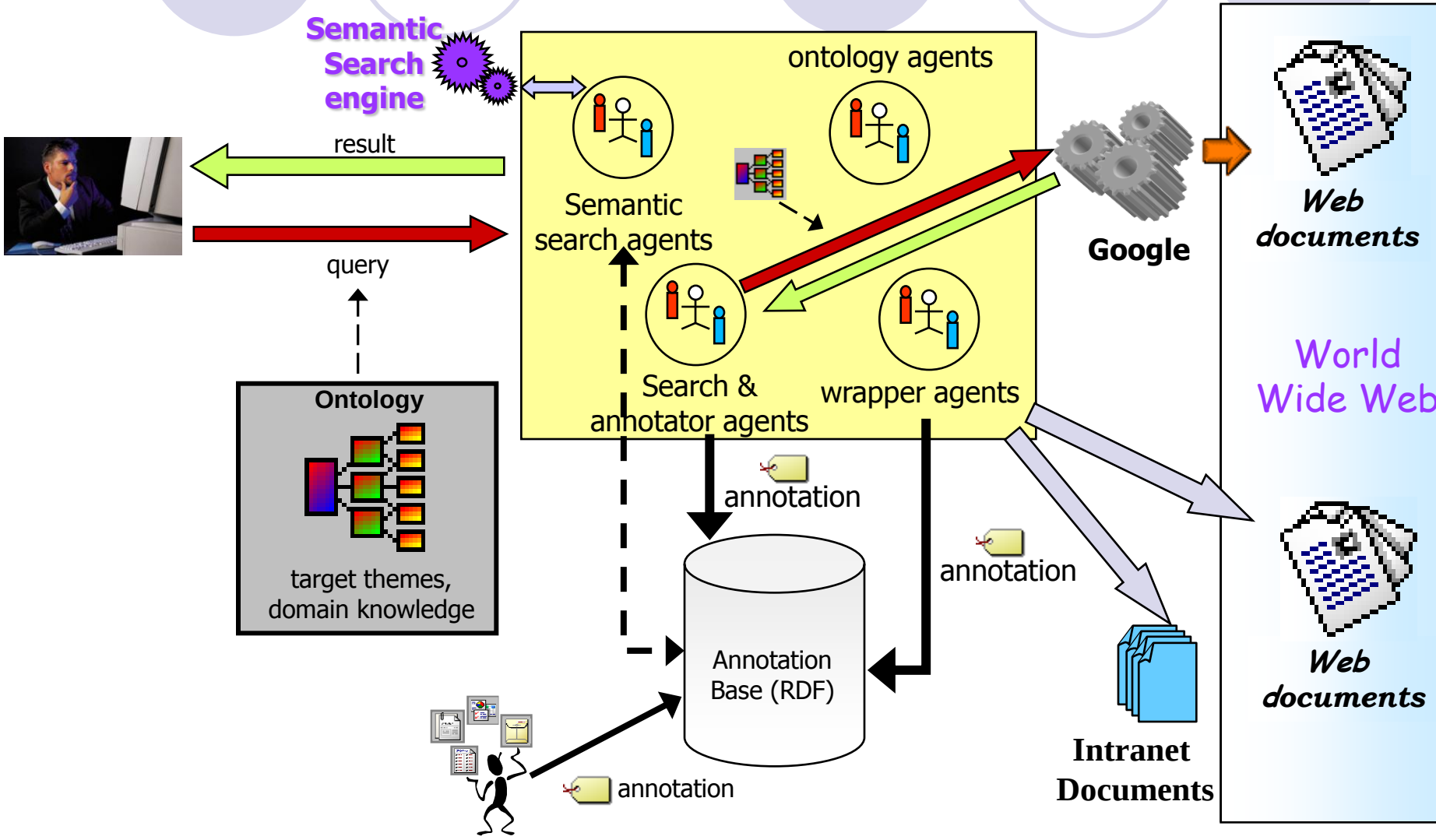
My case: Exploting SW technologies for technology monitoring

- Technology monitoring consists of monitoring the environment of an organization in order to discover the most recent technological and scientific knowledge, to collect and process all the relevant information, likely to make the organization flourishing at short or long term.
- Web is now considered as a hugest information source, need to be exploited in TM.
  - Information about invention, patent, innovation.
  - Technological trends

A Technology monitoring process



Semantic Web approach



Tools for Semantic Web application development

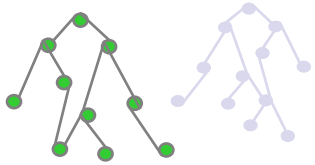
- Create, edit, manage ontology: Protégé 2000.
- Semantic annotation: Melita, MnM, Annotate, Scream,..
- RDF/S engine: Jena, Corese
- RDF/S semantic search engine: Corese

JENA

- Jena is a Java framework for building Semantic Web applications. It provides a programmatic environment for RDF, RDFS and OWL, including a rule-based inference engine.
- Jena is open source and grown out of work with the HP Labs Semantic Web Program.
- The Jena Framework includes:
  - A RDF API
  - Reading and writing RDF in RDF/XML, N3 and N-Triples
  - An OWL API
  - In-memory and persistent storage
  - RDQL – a query language for RDF

http://jena.sourceforge.net/

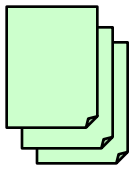
Corese - **Co**nceptual **Re**source **S**earch **E**ngine



Ontologies

```
<rdfs:Class rdf:ID="entity"/>
<rdfs:Class rdf:ID="document">
 <rdfs:subClassOf
    rdf:resource="#entity"/>
</rdfs:Class>
```

RDF Schema



Documents

```
<ns:document rdf:about="http://www.cstb.fr/aar.doc">
  <ns:title>Annual activity report of ACACIA</ns:title>
  <ns:author>
    <ns:person rdf:about="http://www.inria.fr/~dieng"/>
  </ns:author>
</ns:document>
```

RDF Annotations

User

```
?doc rdf:type ns:Document
?doc ns:concern ?topic
?topic rdf:type ns:Java
```

RDF Query

Translation to Conceptual Graph

Searching with capacity of CG

Results in CGs

Result in RDF

CORESE

Functionalities

- Manipulate, validate, resolve queries based on ontologies
- Improve the information retrieval with semantic and reasoning, approximate searching
- Semantic Portal & Server
- Dynamic Interfaces
- Integration XHTML + XML + XSLT + RDF

Query language

SPARQL based query:

select data **where** exp

exp :

resource property value
?x rdf:type c:Person

resource operator value
?name = “Olivier”

«find documents about Java »

?doc rdf:type c:Document
?doc c:concern ?topic
?topic rdf:type c:Java

«find documents about Java and return their title and author»

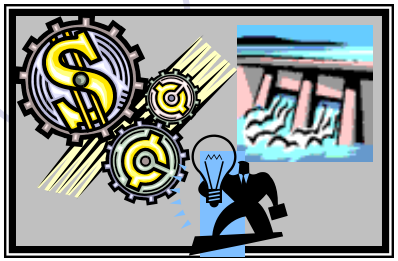
select ?doc ?title ?person **where**
?doc rdf:type c:Document
?doc c:concern ?topic
?topic rdf:type c:Java
?doc c:title ?title
?doc c:author ?person

Inference & Rules

- *If a person wrote a thesis about a topic then he is a doctor and a expert of this topic.*
- ?person author ?doc
- ?doc rdf:type PhDThesis
- ?doc concern ?topic
- $\rightarrow$
- ?person expertIn ?topic
- ?person rdf:type PhD

Thanks!

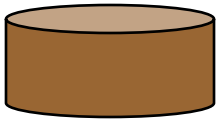
Semantic Web: which resources to annotate ?



Industrial and business processes

External world resources

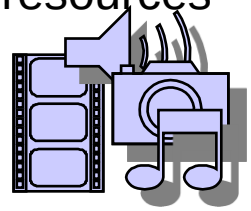
Web resources / services / DBs / etc.



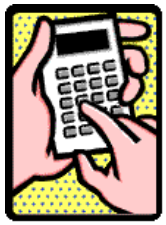
Shared ontology

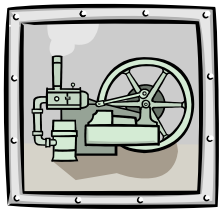
Web users (profiles, preferences)

Multimedia resources



Web access devices





Smart machines and devices

Web agents / applications

semaview.

http://www.semaview.com

The Semantic Web Illustrated

1 It's evolution not revolution.

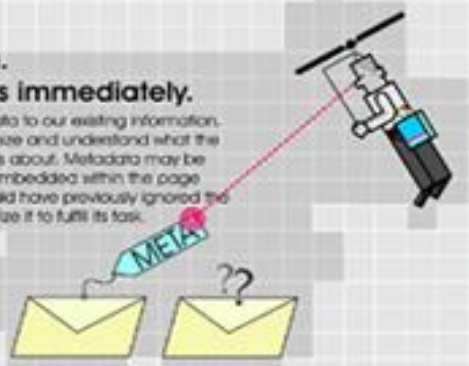
The Semantic Web leverages the content and services that already exist in the World Wide Web. By extending the World Wide Web to be more usable to robots, agents or machines, new applications are now possible.

The Semantic Web is a layer of machine understandable metadata on top of the World Wide Web

The World Wide Web is the foundation for the emerging Semantic Web

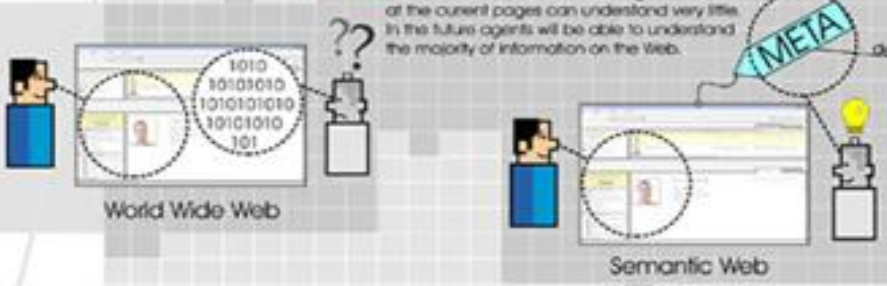
2 Add metadata. Serve to agents immediately.

By attaching additional data to our existing information, agents are able to recognize and understand what the information is and what it is about. Metadata may be linked to a web page or embedded within the page itself. Where an agent would have previously ignored the information, it can now utilize it to fulfill its task.



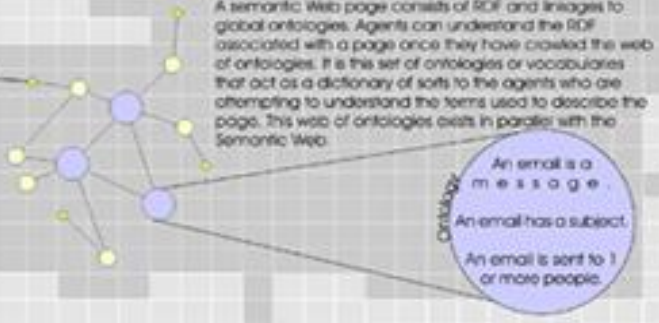
3 What was confusing, now makes perfect sense.

Currently, a Web page is developed mainly for human consumption. Software agents that look at the current pages can understand very little. In the future agents will be able to understand the majority of information on the Web.



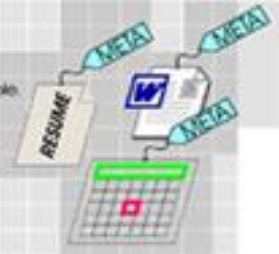
4 Ontologies give the metadata meaning.

A semantic Web page consists of RDF and linkages to global ontologies. Agents can understand the RDF associated with a page once they have crawled the web of ontologies. It is this set of ontologies or vocabularies that act as a dictionary of sorts to the agents who are attempting to understand the terms used to describe the page. This web of ontologies exists in parallel with the Semantic Web.



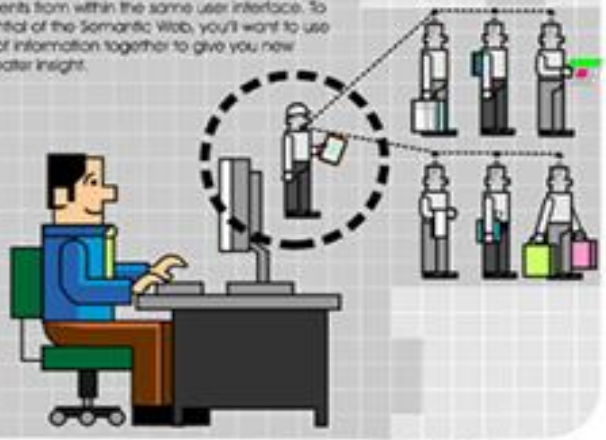
5 The Semantic Web isn't just about Web sites either!

The Semantic Web aims to make many different types of data usable. Emails, Web sites, calendars, resumes, office documents, and contacts are just some of the information resources that can be utilized by software agents if they are enhanced with metadata. Imagine what new applications are possible once agents can tell the difference between these items!



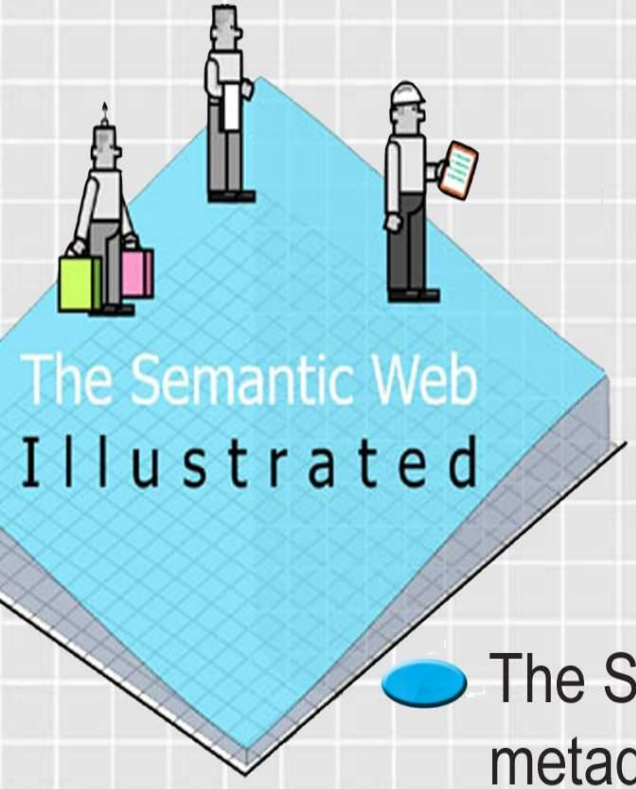
6 How will I use the Semantic Web?

In all likelihood, you'll use the Semantic Web using many different applications. You may have particular applications that use a specific type of information, like calendaring for example. Alternatively, you may choose an application that allows you to control a bevy of agents from within the same user interface. To harness the full potential of the Semantic Web, you'll want to use the different islands of information together to give you new applications and greater insight.



Demonstrating the Value of Technology

1 It's evolution not revolution

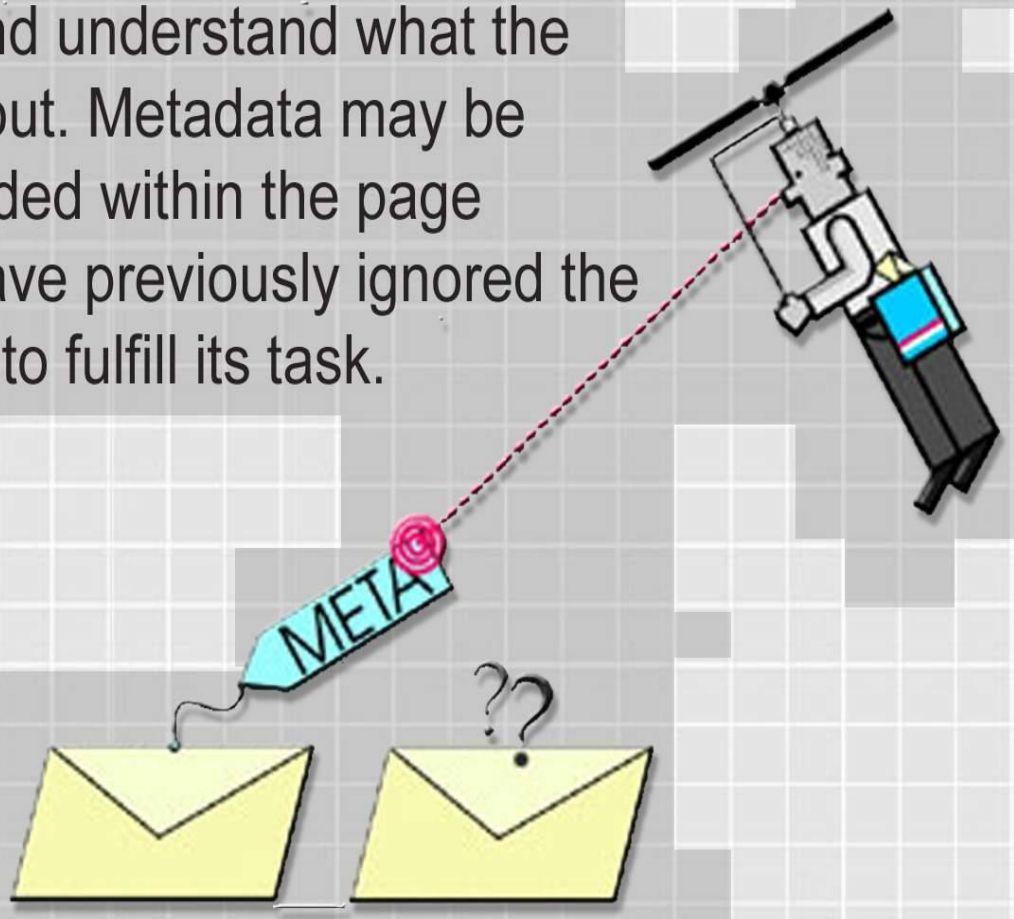


The Semantic Web leverages the content and services that already exist in the World Wide Web. By extending the World Wide Web to be more usable to robots, agents or machines, new applications are now possible.

- The Semantic Web is a layer of machine understandable metadata on top of the World Wide Web
- The World Wide Web is the foundation for the emerging Semantic Web

2 Add metadata. Serve to agents immediately

By attaching additional data to our existing information, agents are able to recognize and understand what the information is and what it is about. Metadata may be linked to a web page or embedded within the page itself. Where an agent would have previously ignored the information, it can now utilize it to fulfill its task.

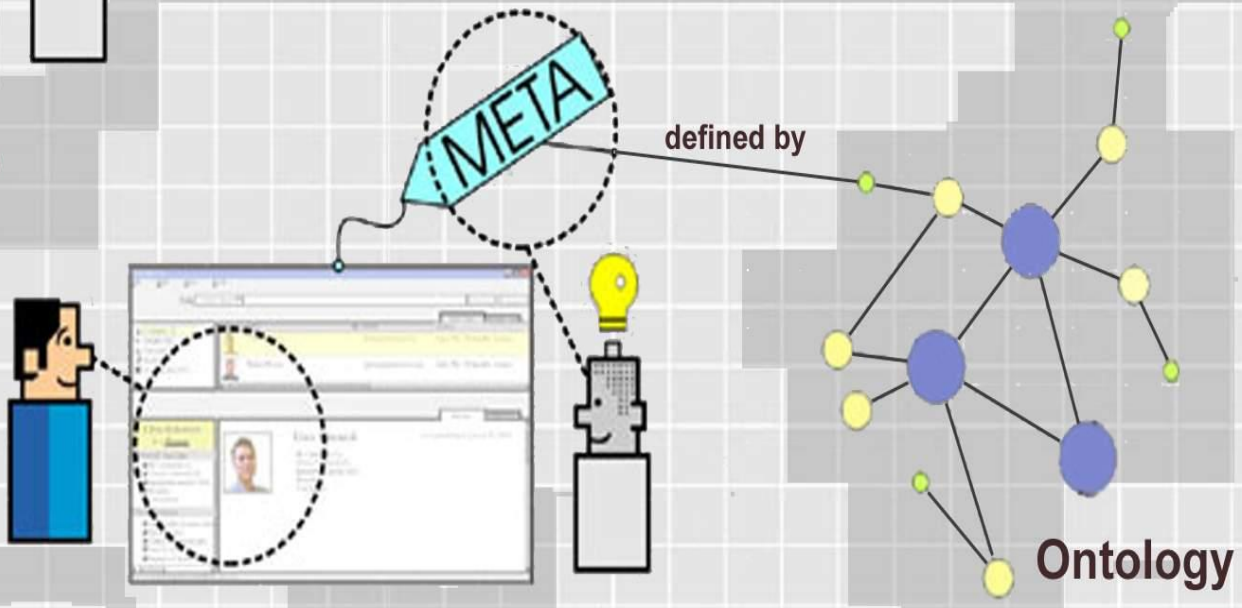


RDF Metadata

3 What was confusing, now makes perfect sense.

Currently, a Web page is developed mainly for human consumption. Software agents that look at the current pages can understand very little. In the future agents will be able to understand the majority of information on the Web.

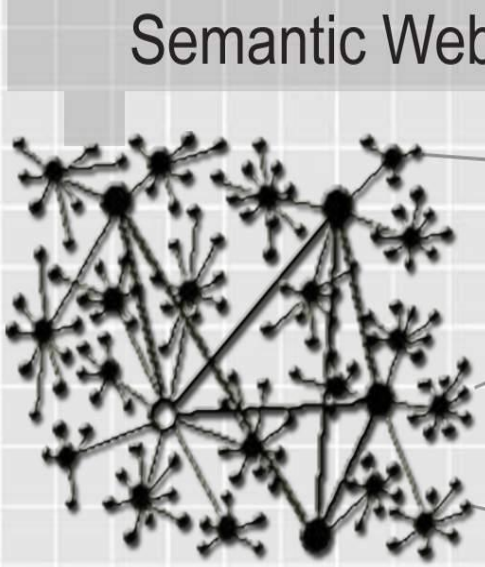
World Wide Web

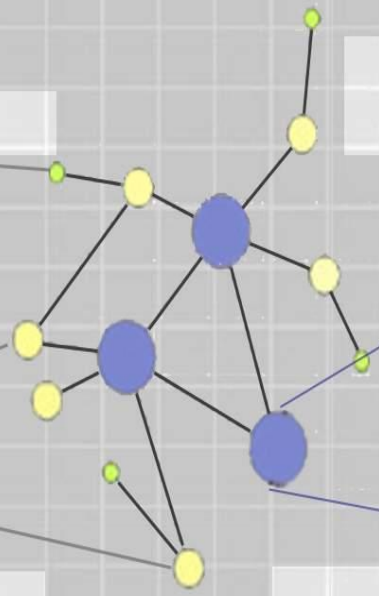


Semantic Web

4 Ontologies give the metadata meaning.

A semantic Web page consists of RDF and linkages to global ontologies. Agents can understand the RDF associated with a page once they have crawled the web of ontologies. It is this set of ontologies or vocabularies that act as a dictionary of sorts to the agents who are attempting to understand the terms used to describe the page. This web of ontologies exists in parallel with the Semantic Web.



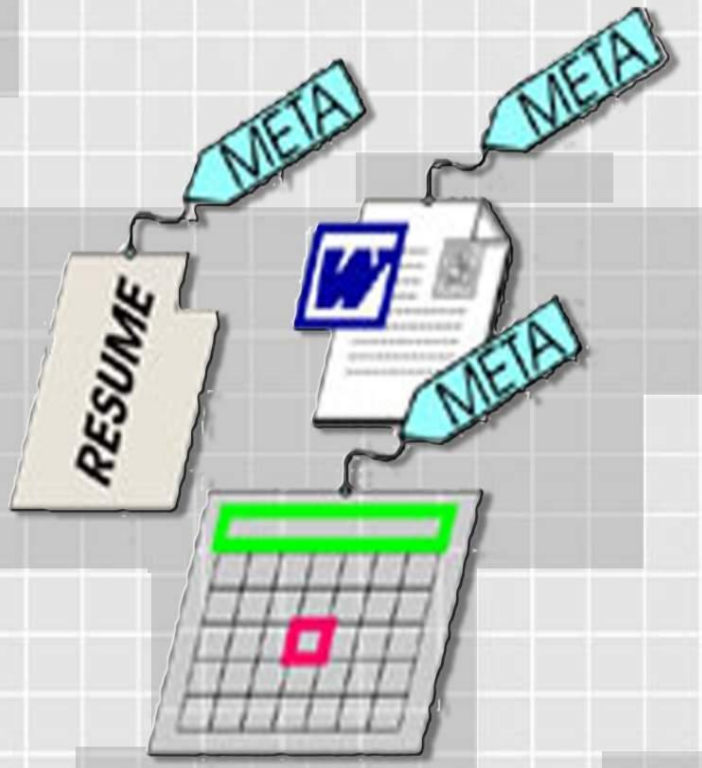


Ontology

An *email* is a *message*
An *email* has a *subject*
... ...
An *email* is sent by a *sender* to 1 or more *reciever*

5 The Semantic Web isn't just about Web site either!

The Semantic Web aims to make many different types of data usable. Emails, Web sites, calendars, resumes, office documents and contacts are just same of the information resources that can be utilized by software agents if they are enhanced with metadata.



Imagine what new application are possible once agents can tell the difference between theses items!

6 How will we use the Semantic Web?

We'll use the Semantic Web using many different applications. We may have particular applications that use a specific type of information, like calendaring for example.

Alternatively, it is possible to choose an application that allows controlling a set of agents from within the same user interface. To harness the full potential of the Semantic Web, the different island of information will be used together to give new applications and greater insight.