Multimedia on the Semantic Web

Jacco van Ossenbruggen, Lynda Hardman
Multimedia and Human-Computer Interaction
CWI Amsterdam

With contributions from
Frank van Harmelen
AI Department
Vrije Universiteit Amsterdam

1

Talk overview

- Three generations of the Web
  - Problems with the current 2nd generation
  - The Semantic Web: a vision of the 3rd generation
- 2nd generation multimedia
  - Cuypers
- Semantic Web technology
  - XML, RDF and DAML+OIL
- Future directions
  - Multimedia on the Semantic Web

2

The Web in three generations

1 Hand-coded (HTML) Web content
easy access through uniform interface
huge authoring and maintenance effort
hard to deal with dynamically changing content

2 Automated on-the fly content generation
based on templates filled with database content
later extended with XML document transformations

3 Automated processing of content
The Semantic Web

3

Who's afraid of the Semantic Web?

- It is not about "blue sky" researchers trying to model the entire world...
- instead, the Semantic Web
  - proposes explicit meta-data rather than "*screen scraping*"
  - by using agreed upon semantics (*ontologies*)
  - building on proven Web technology (XML, RDF, DAML+OIL)

4

Semantic Web application areas

- Search engines
- Browsing on-line stores (B2C)
- Multimedia

5

Problems with current search engines

- Current search engines = keywords:
  - high recall, low precision
  - sensitive to vocabulary
  - insensitive to implicit content

6

Search engines on the Semantic Web

- concept search instead of keyword search
- semantic narrowing/widening of queries
- query-answering over >1 document
- document transformation operators

7

Problems with 2nd generation on-line stores (B2C)

- manual browsing is time-consuming and inefficient
- every shopbot requires a series of wrappers
  - work only partially
  - extract only explicit information
  - must be updated frequently

8

B2C on the Semantic Web

- Software agents “understand” product descriptions
  - enabling automatic browsing
- Procedural wrapper-coding becomes declarative ontology-mapping
  - improving robustness and simplifying maintenance

9

Multimedia scenario

User is taking an art class on Rembrandt and wants to know about the “*chiaroscuro*” technique

System responds with a textual and audio explanation of the technique and a number of example images of its application in Rembrandt’s paintings

10

2nd generation multimedia

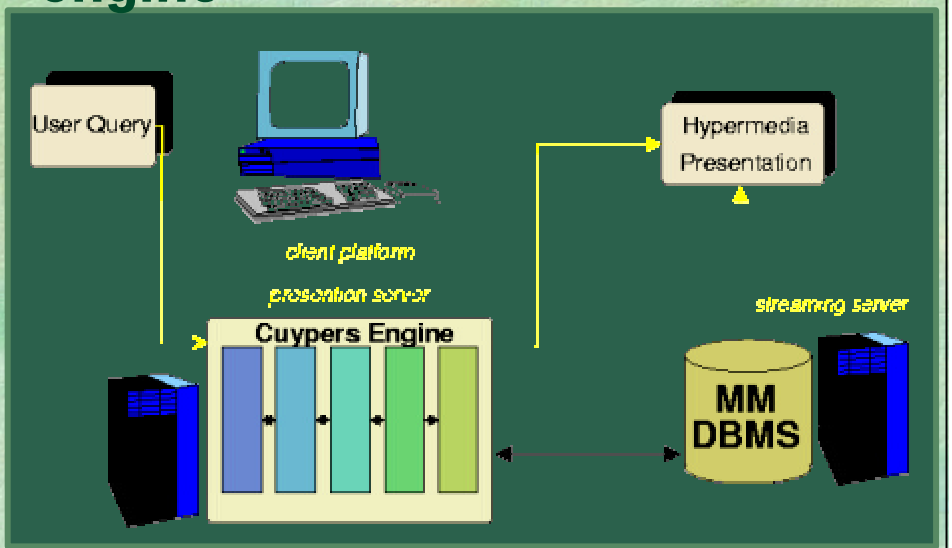
- Adapt to end-user's platform capabilities
  - PC, PDA, mobile, voice-only, …
- Adapt to the network resources available
  - bandwidth and other quality of service parameters
- Personalization
  - language, abilities, level of expertise, …

- Problem: current 2nd generation Web tools

  do not work for multimedia

11

Cuypers multimedia generation engine



12

Cuypers multimedia generation engine

- Demo time

- **Acknowledgements:**
  - **Demonstrator developed in the context of the ToKeN2000 project**
  - **Media database used with permission, courtesy Rijksmuseum Amsterdam.**

13

Cuypers – the bad news

Currently all our design knowledge is:

- implicit and hidden in the generation rules
- lost in the generated Web presentation
- not reusable for other Web applications/sites

We need the Semantic Web

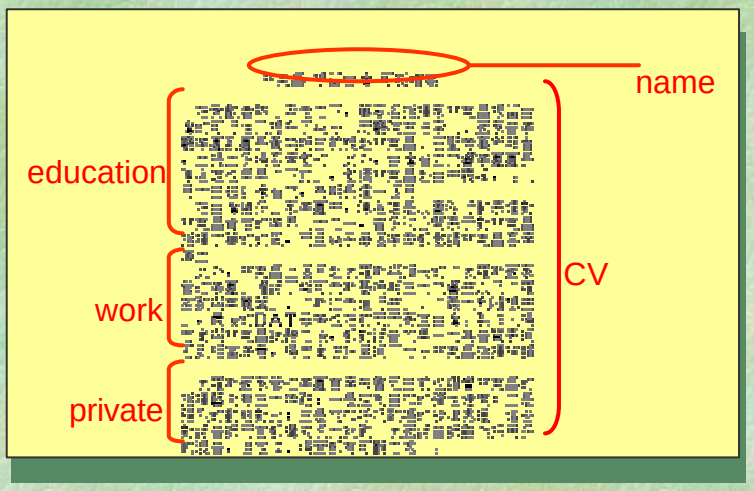
14

So what *is* the Semantic Web?

15

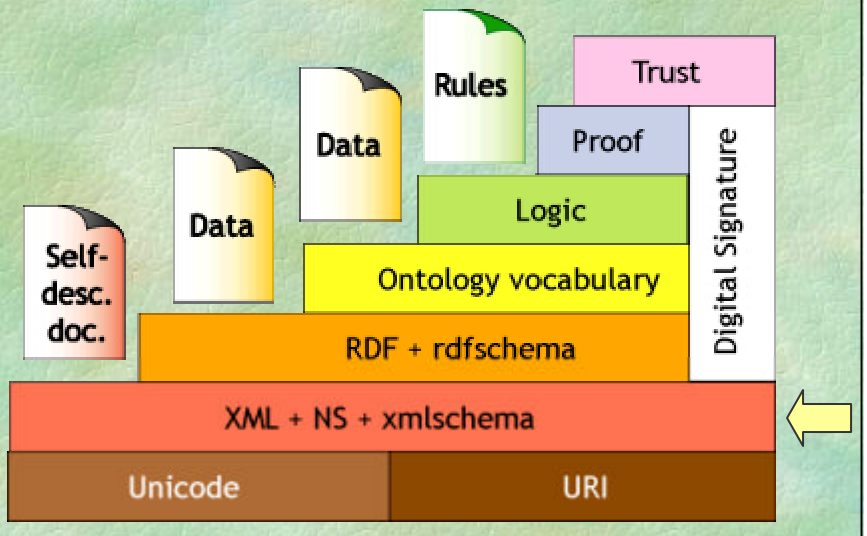
Machine accessible meaning

(What it's like to be a machine)



16

TBL talk at XML 2000



XML: User definable and domain specific markup

HTML:

```
<H1>Introduction to AI</H1>
     <UL> <LI>Teacher: Frank van Harmelen
          <LI>Students: 1AI, 1I
          <LI>Requirements: none
     </UL>
```

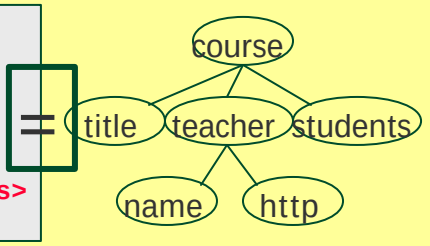
XML:

```
<course>
     <title>Introduction to AI</title>
     <teacher>Frank van Harmelen</teacher>
     <students>1AI, 1I</students>
     <req>none</req>
</course>
```

XML: document = labelled tree

- node = label + attr/values + contents

```
<course date="...">
      <title>...</title>
      <teacher>...</teacher>
            <name>...</name>
            <http>...</http>
      <students>...</students>
</course>
```



- schema: simple grammars to describe legal trees
- So:

why not use XML to represent ontologies?

19

XML: limitations for semantic markup

XML makes no commitment on:

- ✍ Domain-specific ontological **vocabulary**
- ✍ Ontological **modeling primitives**

✍ requires pre-arranged agreement on ✍ & ✍

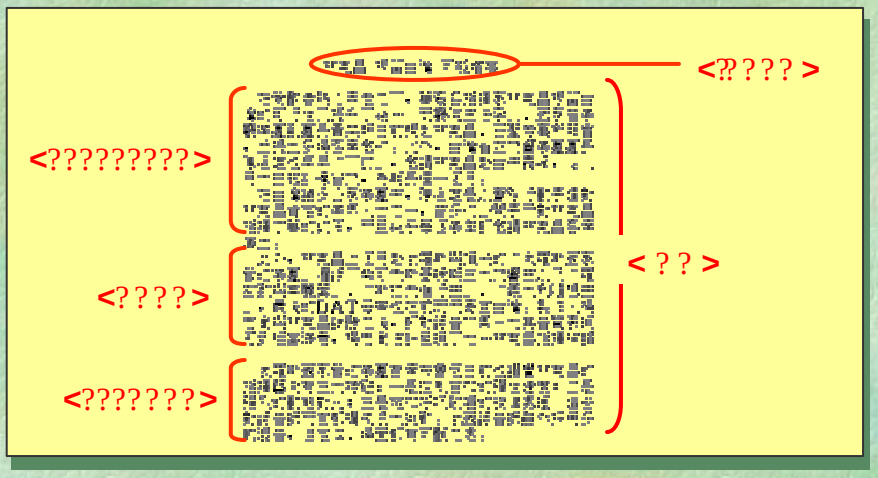
Only feasible for closed collaboration

- agents in a small & stable community
- pages on a small & stable intranet

not for sharable Web-resources ☹

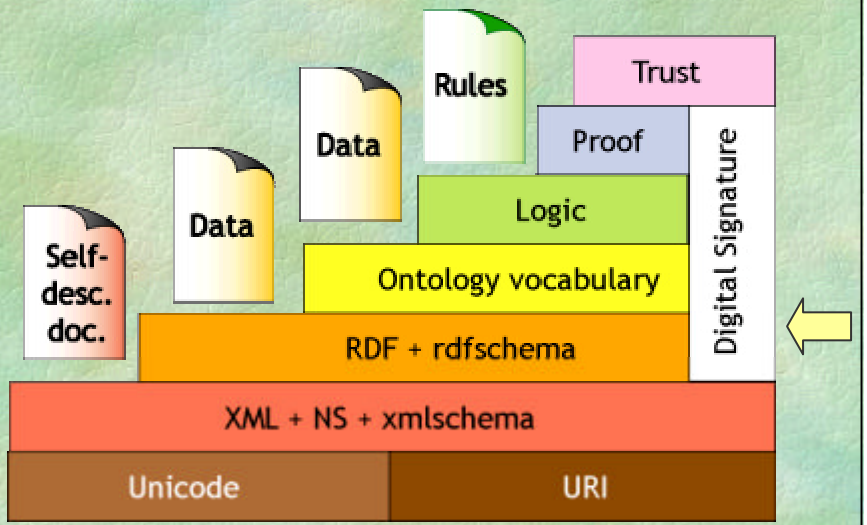
20

XML ?
machine accessible meaning



21

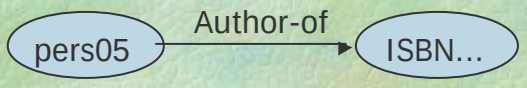
The semantic pyramid again



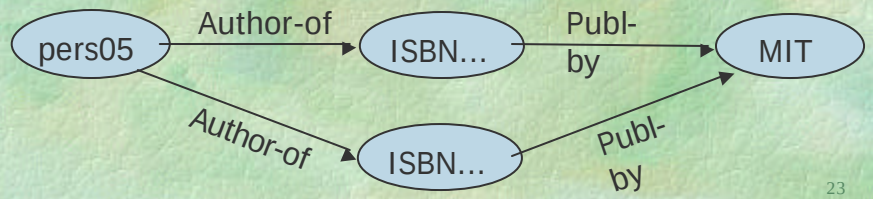
22

RDF: graphs of triples

- Object ->Attribute-> Value triples



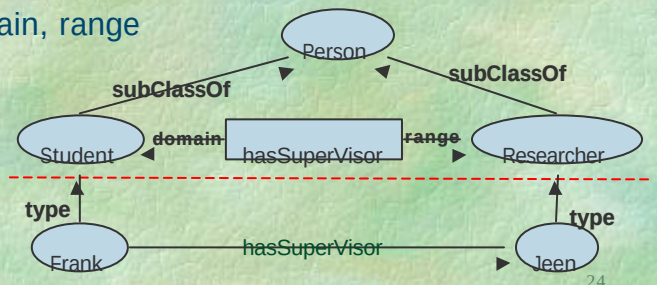
- Objects are web-resources
- Value is again an Object:
  - triples can be linked
  - data-model = graph



23

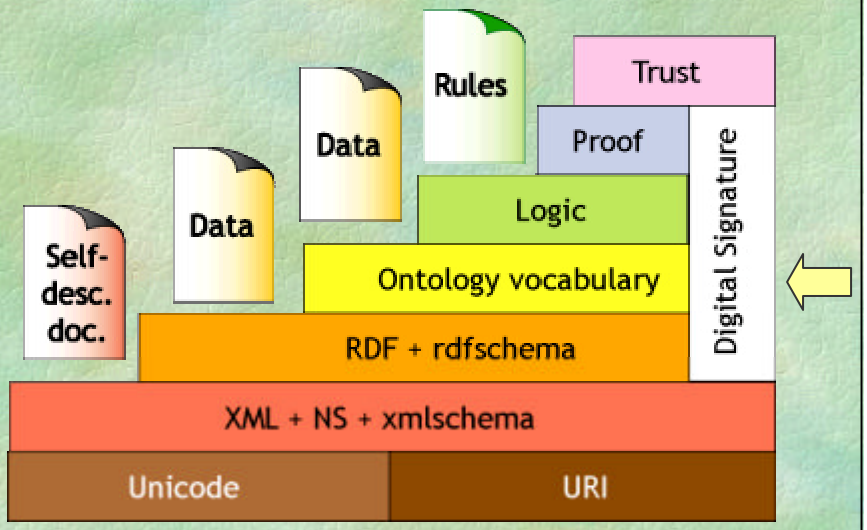
What does RDF Schema add?

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



24

The semantic pyramid again



25

WebOnt and OntoWeb

- W3C WebOnt working group set up 1 Nov 2001
  Work continuing where DAML+OIL left off
  http://www.w3.org/2001/sw/WebOnt/charter
- WebOnt is part of W3C Semantic Web activity which also includes RDF
- OntoWeb
  EU funded thematic network
  > 80 partners, including CWI and VU
  http://www.ontoweb.org

26

Semantic Web: main players

Academic in Europe:

- VU, Amsterdam
- Karlsruhe
- Manchester
- INRIA
- SWI@UvA

Academic in US:

- Stanford
- Maryland
- MIT/W3C
- Florida
- CMU

Industrial:

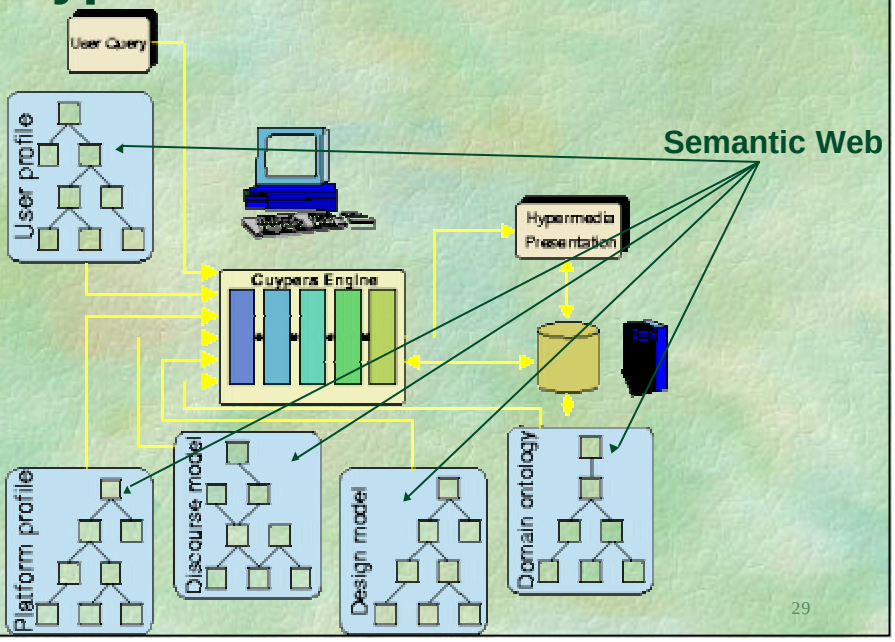
- Lucent
- Philips
- Nokia
- HP
- Intel
- Daimler-Chrysler
- Fujitsu
- lots of start-ups (NL, UK, G, N, US)

SW isn't just KR in XML/RDF

- It's large
- It's even larger
- no referential integrity
- many authors, distributed authority, trust
- high variety in quality of knowledge
- diverse vocabularies
- decentralized
- high change rate, time-dependent content
- local containment of inconsistencies
- justifications as first order citizens

28

Cuypers revisited



29

Conclusions

XML technology is commonplace, but

- insufficient for multimedia generation
  - CWI's Cuypers realises 2nd generation multimedia
- insufficient for machine understandable metadata
  - RDF(S) provides basic KR primitives
  - WebOnt is developing W3C ontology language

3rd generation MM focus of current research

- reusing knowledge available on the Semantic Web
- generating annotated multimedia

30