Semantic Web

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

With contributions from
Frank van Harmelen, Heiner Stuckenschmidt
AI Department
Vrije Universiteit Amsterdam

Talk overview

- **Three generations of the Web**
  - Problems with the current 2nd generation
  - The Semantic Web: a vision of the 3rd generation
- Semantic Web technology
  - XML, RDF
  - DAML+OIL tutorial
- 2nd generation multimedia
  - Cuypers
- Towards 3rd generation multimedia

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 (SW)

SW application areas

- Search engines
- Browsing on-line stores (B2C)
- Service description and integration (B2B)
- Tailored multimedia information

Problems with current search engines

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

Image search on the Web

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

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

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

Current B2B problems

- EDIFACT

EDIFACT S93A Sample Document

PURCHASE ORDER

```
UNB+UNOB:1+003897733:01:MFGB-PO+PARTNER ID:ZZ+000101:1050
+00000000000916++ORDERS'
UNH+1+ORDERS:S:93A:UN'
BGM+221+P1M24987E+9'
DTM+4:20000101:102'
FTX+PUR+3++PURCHASE ORDER BEFORE LINE ITEM INSTRUCTIONS'
RFF+CT:123-456'
NAD+SE+10025392::92++SUPPLIER NAME'
CTA+SR+:STEVE'
NAD+BT+B2::92++COMPAQ COMPUTER CORPORATION+P O BOX 692000
+HOUSTON+TX+77692000+US'
NAD+BY+MFUS::92++COMPAQ COMPUTER CORPORATION'
CTA+PD+:CLARETTA STRICKLAND-FULTON'
NAD+ST+CM6::92++COMPAQ COMPUTER CORPORATION+CCM6 RECEIVING DOCK:20555 SH
249+HOUSTON+TX+77070+US'
TAX+9++++++3-00105-5135-3'
CUX+2:USD:9'
PAT+1++1:1:D:45'
PCD+12:2'
TDT+20++++:::AIRBORNE'
LOC+16+COMPAQ DOCK'
TOD+2+NS+:::ORIGIN COLLECT'
IA+1+AA:EC+123456:VP'
IMD+F+8+:::PART DESCRIPTION INFORMATION
```

B2B on the Semantic Web

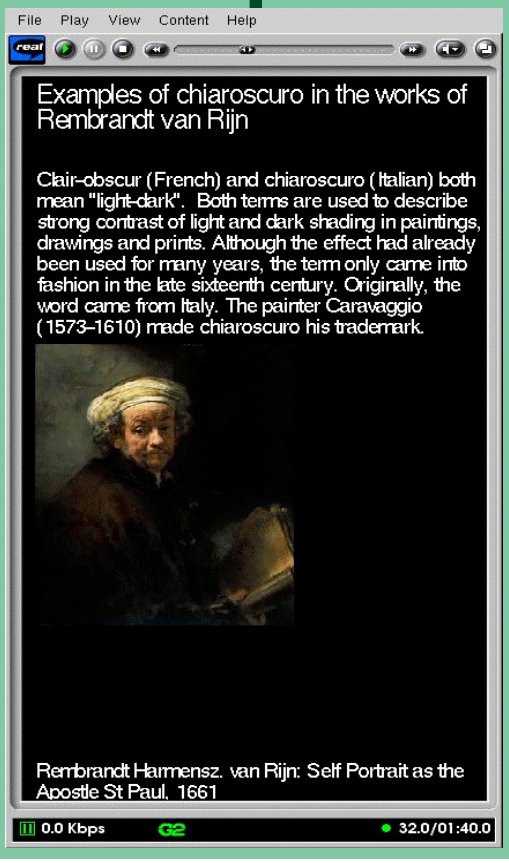
- Semantic Web technology is
  - Cheaper
  - Flexible
  - Integrated with “document” Web
- Provides interoperable semantics for
  - vertical markets (verticalnet.com)
  - horizontal markets

VerticalNet

OASIS

RosettaNet

Example scenario



Student 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

Accessing information on the 2nd generation Web

- Students have access to material on the Web
  - Search problem
- Material is designed for “typical” student
  - No student is typical
- Some adaptivity is possible
  - Links revealed once material has been covered
- Student’s knowledge level is implicit

Accessing information on the Semantic Web

- Students would be able to find suitable courses
- Material can be tailored for the individual
- Material can be re-used
- Models can be made of
  - The domain
  - Learner profile
  - Learning strategies
- Student’s knowledge level can be made explicit
  - in terms of the domain model
  - in terms of the learning strategy

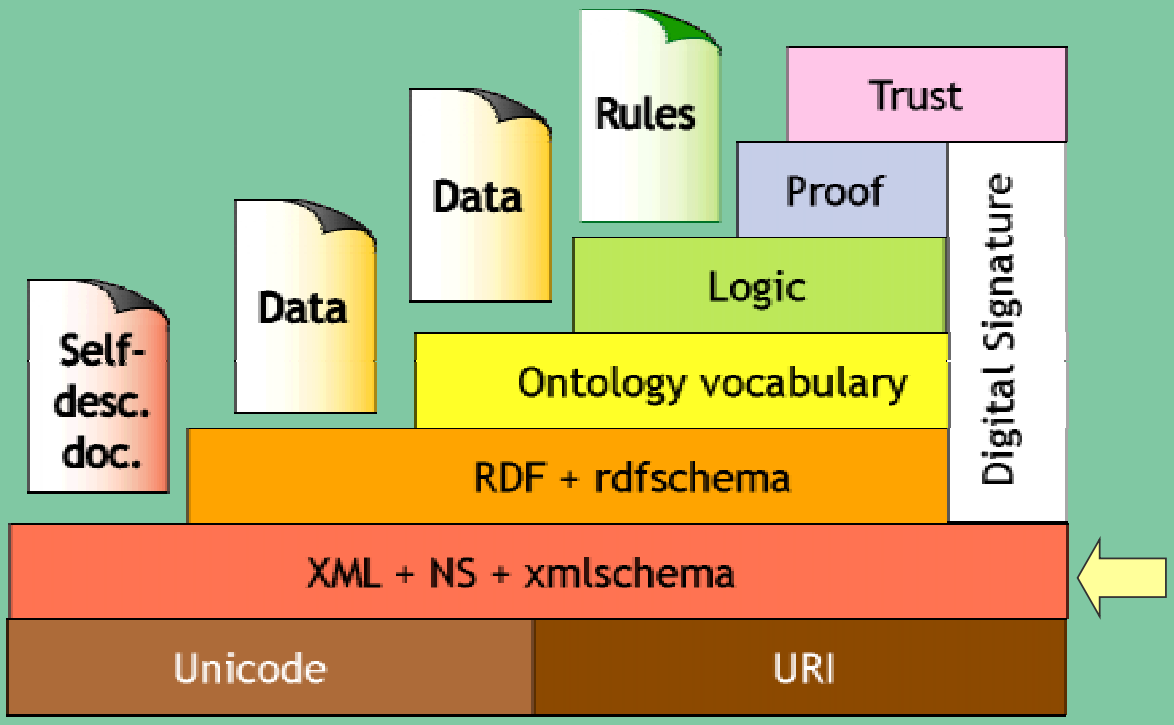
Talk Overview

- Three generations of the Web
  - Problems with the current 2nd generation
  - The Semantic Web: a vision of the 3rd generation
- **Semantic Web technology**
  - XML, RDF
  - DAML+OIL tutorial
- 2nd generation multimedia
  - Cuypers
- Towards 3rd generation multimedia

So what is the Semantic Web?

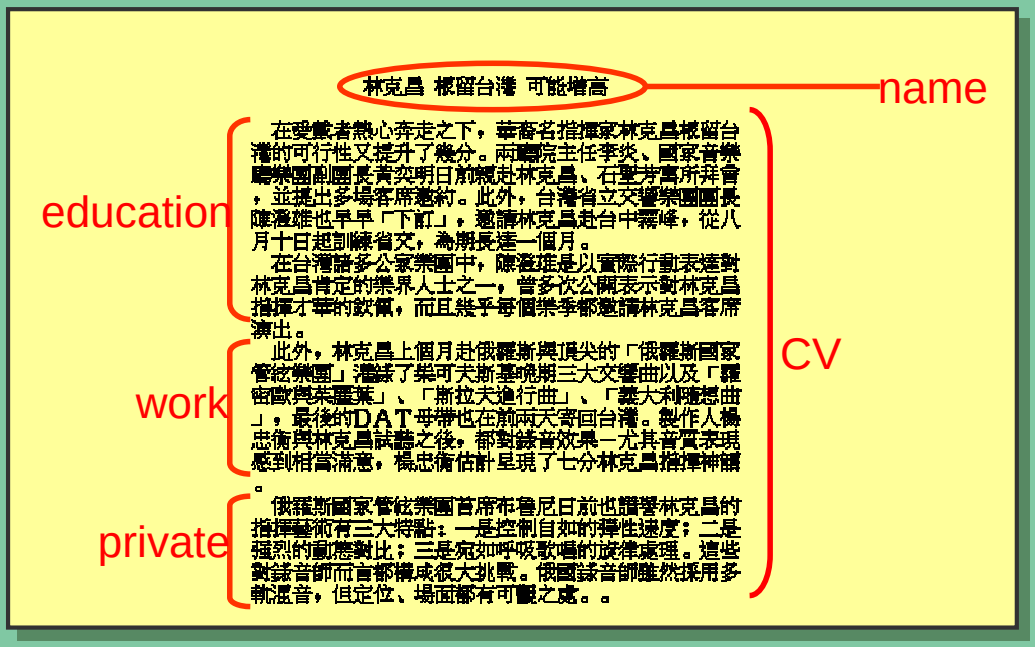
- The “Semantic Web Wedding Cake”
- Crash course
  - XML
  - RDF
  - RDF Schema
  - DAML+OIL (OWL)

TBL talk at XML 2000



Machine accessible meaning

(What it's like to be a machine)



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:

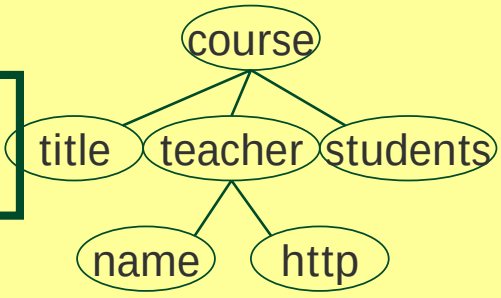
```
<course>
     <title>Introduction to AI</title>
     <teacher>Frank van Harmelen</teacher>
     <students>1AI, 1I</students>
     <req>none</req>
</course>
```

XML: document = labelled tree

- node = label + attr/values + contents

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

=



- XML Schema: grammars for describing legal trees and datatypes
- So:

why not use XML to represent semantics?

XML: limitations for semantic markup

- XML makes no commitment on:
  - ❶ Domain-specific ontological vocabulary
  - ❷ Ontological modeling primitives
- Requires pre-arranged agreement on ❶ & ❷
- Only feasible for closed collaboration
  - agents in a small & stable community
  - pages on a small & stable intranet
- Not suited for sharing Web-resources

XML ≠ machine accessible meaning

林克昌 根留台灣 可能增高 < ναμε >

<εδυχατιον>

在愛樂者熱心奔走之下，華裔名指揮家林克昌根留台灣的可行性又提升了幾分。兩廳院主任李炎、國家音樂廳樂團副團長黃奕明日前就赴林克昌、石聖芳高所拜會，並提出多場客席邀約。此外，台灣省立交響樂團團長陳澄雄也早早「下訂」，邀請林克昌赴台中霧峰，從八月十日起訓練省交，為期長達一個月。

在台灣諸多公家樂團中，陳澄雄是以實際行動表達對林克昌肯定的樂界人士之一，曾多次公開表示對林克昌指揮才華的欽佩，而且幾乎每個樂季都邀請林克昌客席演出。

<ωορκ>

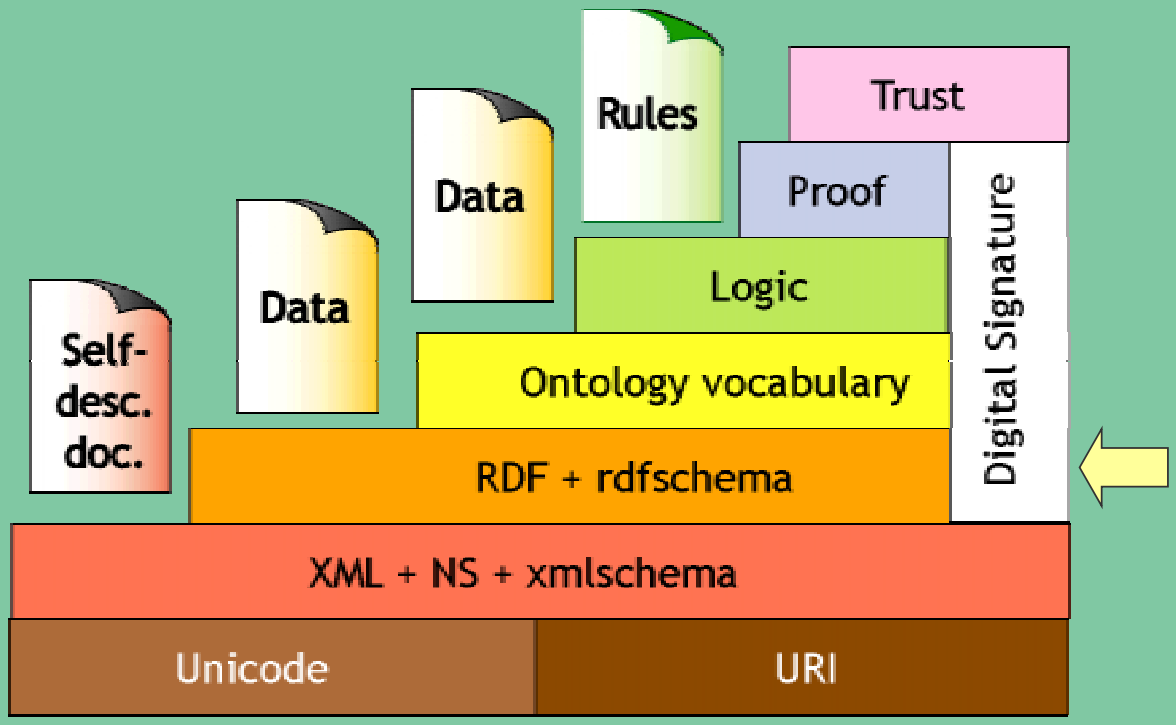
此外，林克昌上個月赴俄羅斯與頂尖的「俄羅斯國家管絃樂團」灌錄了柴可夫斯基曉期三大交響曲以及「羅密歐與茱麗葉」、「斯拉夫進行曲」、「義大利隨想曲」，最後的DAT母帶也在前兩天寄回台灣。製作人楊忠衡與林克昌試聽之後，都對錄音效果一尤其音質表現感到相當滿意，楊忠衡估計呈現了七分林克昌指揮神韻。

< Χς >

<πριωατε>

俄羅斯國家管絃樂團首席布魯尼日前也讚譽林克昌的指揮藝術有三大特點：一是控制自如的彈性速度；二是強烈的動態對比；三是宛如呼吸歌唱的旋律處理。這些對錄音師而言都構成很大挑戰。俄國錄音師雖然採用多軌混音，但定位、場面都有可觀之處。。

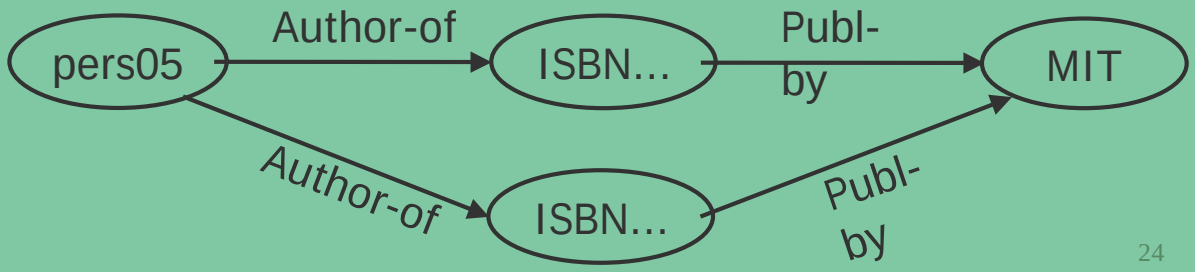
The semantic pyramid again



RDF: graphs of triples

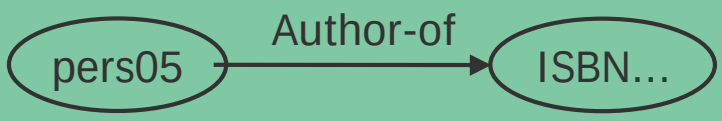
- Object ->Attribute-> Value triples

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



RDF is part of the Web

- Every resource has a URI
  = world-wide unique naming!

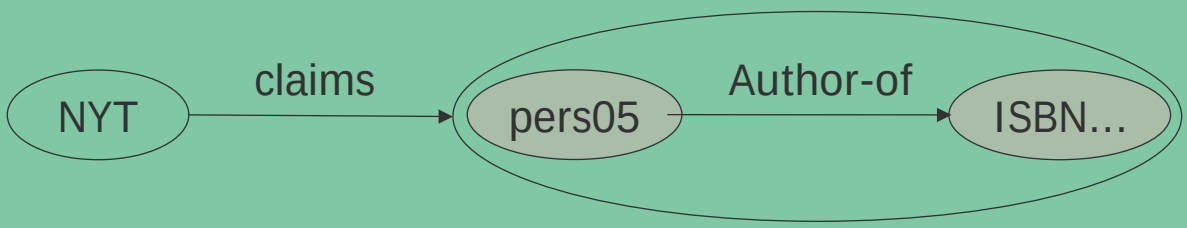


- Has XML syntax(es)

```
<rdf:Description rdf:about="#pers05">
  <authorOf>ISBN...</authorOf>
</rdf:Description>
```


RDF can be nested

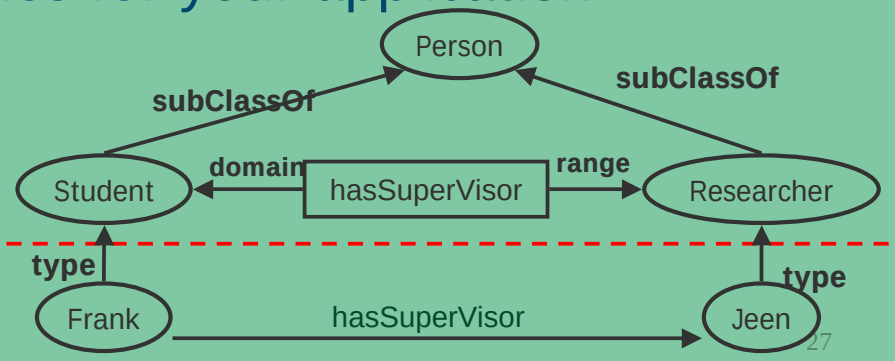
- Any statement can be an object
  graphs can be nested - reification




```
<rdf:Description rdf:about="#NYT">
  <claims>
    <rdf:Description rdf:about="#pers05">
      <authorOf>ISBN...</authorOf>
    </rdf:Description>
  </claims>
</rdf:Description>
```

What does RDF Schema add?

- Defines small **vocabulary** for RDF:
  - Class, subClassOf, type
  - Property, subPropertyOf
  - domain, range
- Vocabulary can be used to define other vocabularies for your application domain



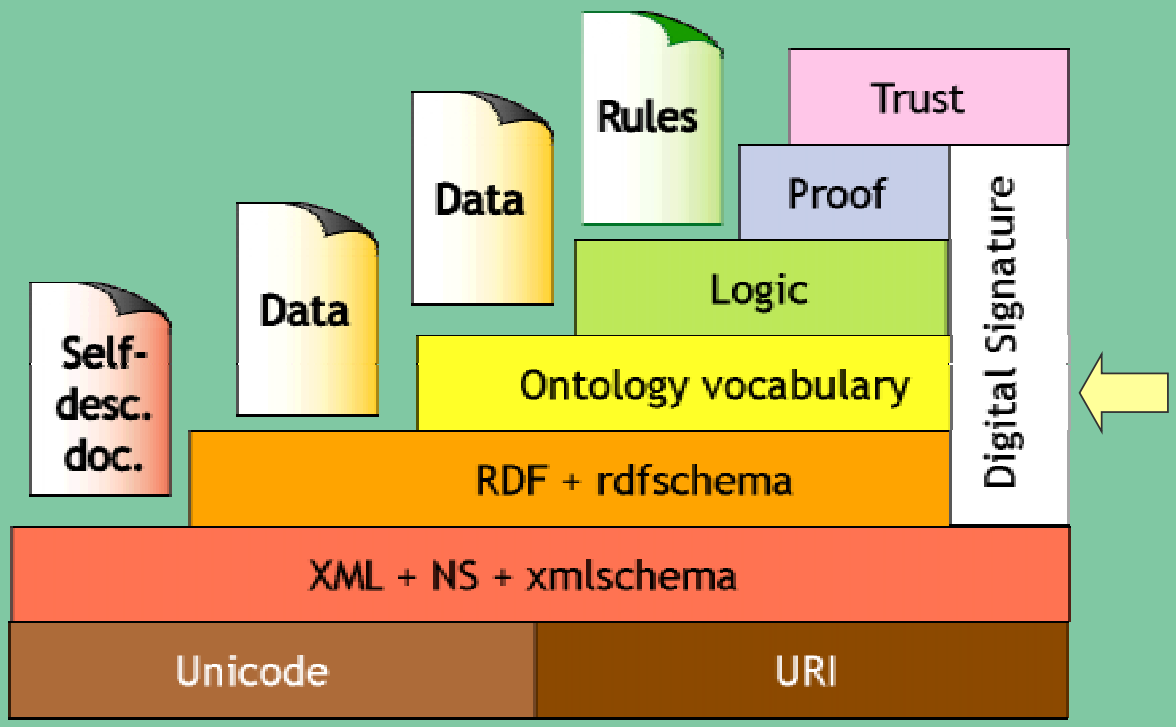
RDF Schema syntax in XML

```
<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>
```

Conclusions about RDF(S)

- Next step up from plain XML ?
  - (small) ontological commitment to modeling primitives
  - possible to define vocabulary
- However:
  - no precisely described meaning
  - no inference model

The semantic pyramid again



Beyond RDF: OIL & DAML

- OIL extends RDF Schema to a fully-fledged knowledge representation language.
  - logical expressions
  - data-typing
  - cardinality
  - quantifiers
  - http://www.ontoknowledge.org
- DAML = US sister of OIL
- Merged as DAML+OIL in 2001
- Becomes OWL W3C standard in March '03

WebOnt and OntoWeb

- W3C WebOnt working group Nov 2001–Mar 2003
  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, UvA and VU
  http://www.ontoweb.org

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)

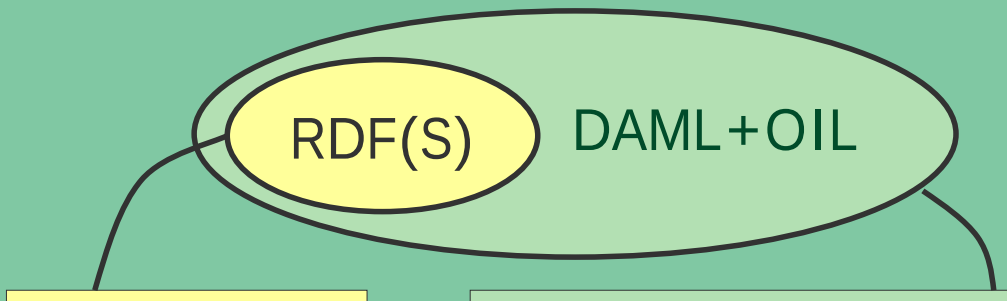
DAML+OIL (by example)

```
class-def animal                          % animals are a class
class-def plant                           % plants are a class
   subclass-of NOT animal                 % that is disjoint from animals
class-def tree
   subclass-of plant                      % trees are a type of plants
class-def branch
   slot-constraint is-part-of             % branches are parts of some tree
      has-value tree
      max-cardinality 1
class-def defined carnivore               % carnivores are animals
   subclass-of animal
   slot-constraint eats                   % that eat any other animals
      value-type animal
class-def defined herbivore               % herbivores are animals
   subclass-of animal, NOT carnivore      % that are not carnivores, and
   slot-constraint eats                   % they eat plants or parts of plants
      value-type plant OR (slot-constraint is-part-of has-value plant)
```

DAML+OIL as RDF(S) extension

```
<rdfs:Class rdf:ID="herbivore">
   <rdf:type
        rdf:resource="http://www.ontoknowledge.org/#DefinedClass"/>
   <rdfs:subClassOf rdf:resource="#animal"/>
   <rdfs:subClassOf>



   </rdfs:subClassOf>
</rdfs:Class>
```


DAML+OIL as RDFS extension



- class-def
- subclass-of
- slot-def
- subslot-of
- domain
- range

- class-expressions
  - AND, OR, NOT
- slot-constraints
  - has-value, value-type
  - cardinality
- slot-properties
  - trans, symm

DAML+OIL: Classes

```
<daml:Class rdf:ID="Male">
     <rdfs:subClassOf rdf:resource="#Animal"/>
</daml:Class>

<daml:Class rdf:ID="Female">
     <rdfs:subClassOf rdf:resource="#Animal"/>
     <daml:disjointWith rdf:resource="#Male"/>
</daml:Class>

<daml:Class rdf:ID="Man">
  ^!  <rdfs:subClassOf rdf:resource="#Person"/>
     <rdfs:subClassOf rdf:resource="#Male"/>
</daml:Class>
```


Class-Building Operations

- Relation to other Classes
  - `rdfs:subClassOf`
  - `daml:disjointWith`
  - `daml:disjointUnionOf`
  - `daml:sameClassAs`
  - `daml:equivalentTo`
- Contained Elements:
  - `daml:oneOf`
- Boolean combinations:
  - `daml:intersectionOf`
  - `daml:unionOf`
  - `daml:complementOf`

DAML+OIL: Properties

```
<daml:ObjectProperty rdf:ID="hasParent">
     <rdfs:domain rdf:resource="#Animal"/>
     <rdfs:range rdf:resource="#Animal"/>
</daml:ObjectProperty>

<daml:ObjectProperty rdf:ID="hasChild">
     <daml:inverseOf
         rdf:resource="#hasParent"/>
</daml:ObjectProperty>

<daml:UniqueProperty rdf:ID="hasMother">
     <rdfs:subPropertyOf
         rdf:resource="#hasParent"/>
     <rdfs:range rdf:resource="#Female"/>
</daml:UniqueProperty>
```

Property-Building Operations

- Basic Types
  - `daml:ObjectProperty`
  - `daml:DataTypeProperty`
- Special Types
  - `daml:TransitiveProperty`
  - `daml:UniqueProperty`
  - `daml:UnambigousProperty`
- Further Restrictions
  - `rdfs:subPropertyOf`
  - `rdfs:domain`
  - `rdfs:range`
  - `daml:samePropertyAs`
  - `daml:inverseOf`

DAML+OIL: Property Restrictions

```
<daml:Class rdf:ID="Person">
  <rdfs:subClassOf rdf:resource="#Animal"/>
  <rdfs:subClassOf>
    <daml:Restriction>
      <daml:onProperty
          rdf:resource="#hasFather"/>
      <daml:toClass rdf:resource="#Man"/>
    </daml:Restriction>
    <daml:Restriction daml:cardinality="1">
      <daml:onProperty
          rdf:resource="#hasFather"/>
    </daml:Restriction>
  </rdfs:subClassOf>
</daml:Class>
```

DAML+OIL: Property Restrictions

```
<daml:Class rdf:ID="Person">
   <rdfs:subClassOf rdf:resource="#Animal"/>
   <rdfs:subClassOf>
      <daml:Restriction daml:cardinalityQ="1">
        <daml:onProperty
            rdf:resource="#hasFather"/>
        <daml:hasClassQ rdf:resource="#Man"/>
      </daml:Restriction>
   </rdfs:subClassOf>
</daml:Class>
```

Restrictions

- General
  - daml:Restriction
  - daml:onProperty
- Number Restrictions
  - daml:cardinality
  - daml:maxCardinality
  - daml:minCardinality
- Value and Type Restrictions
  - daml:toClass
  - daml:hasValue
  - daml:hasClass
- Combinations
  - daml:cardinalityQ
  - daml:maxCardinalityQ
  - daml:minCardinalityQ

DAML+OIL: Individuals

```
<rdf:Description rdf:ID="Asia">
   <rdf:type>
      <rdfs:Class rdf:about="#continent"/>
   </rdf:type>
   <daml:differnetIndividualFrom
      rdf:resource="#Europe"/>
</rdf:Description>

<continent rdf:ID="Asia"/>

<rdf:Description rdf:ID="India">
   <is_part_of rdf:resource="#Asia"/>
   <daml:sameIndividual
      rdf:resource="#IndianSubcontinent"/>
   <inhabitants>
      <xsd:integer rdf:value="700.000.000"/>
   </inhabitants>
</rdf:Description>
```

DAML+OIL: Defined Datatypes

```
<xsd:simpleType name="over17">
   <xsd:restriction
      base="xsd:positiveInteger">
      <xsd:minInclusive value="18"/>
   </xsd:restriction>
</xsd:simpleType>

<daml:Class rdf:ID="Adult">
   <daml:intersectionOf
      rdf:parseType="daml:collection">
      <daml:Class rdf:about="#Person"/>
      <daml:Restriction>
         <daml:onProperty
            rdf:resource="#age"/>
         <daml:hasClass
            rdf:resource="#over17"/>
      </daml:Restriction>
   </daml:intersectionOf>
</daml:Class>
```

DAML+OIL: more info.

- DAML+OIL home page
- Language tutorial
- OWL

SW isn't just KR in XML/RDF

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

Talk overview

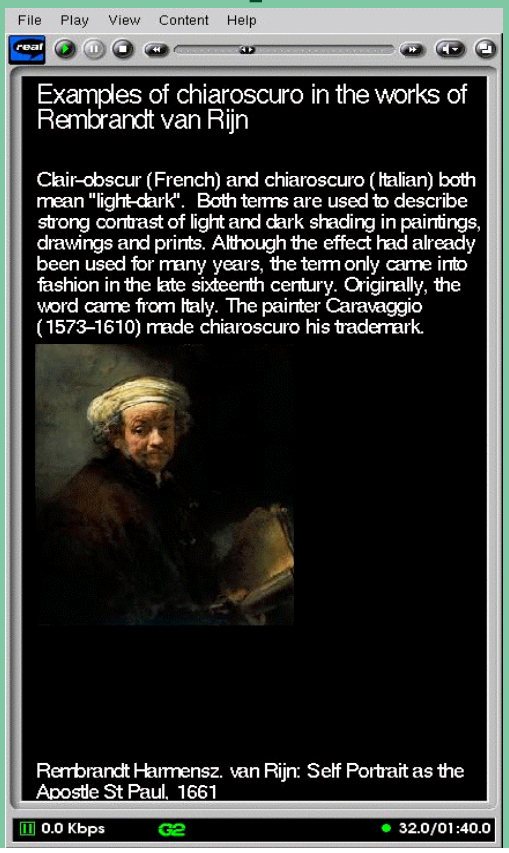
- Three generations of the Web
  - Problems with the current 2nd generation
  - The Semantic Web: a vision of the 3rd generation
- Semantic Web technology
  - XML, RDF and DAML+OIL tutorial
- **2nd generation multimedia**
  - **Cuypers**
- Towards 3rd generation multimedia

Multimedia on the Web

- Real multimedia Web content is still rare
  - Mostly bells & whistles to enhance HTML text ...
  - ... or mono-media AV-streams
- Virtually all presentations are hand-authored
  - proprietary formats that are hard to generate
  - limited support for dynamic content and multichanneling
  - most Web technology is text/page-oriented ...
  - ... with SMIL as one of the few exceptions
- Conclusion:

  Multimedia has hardly caught up with the 1st generation Web!

Example scenario



- User is interested in Rembrandt and wants to know about about the “chiaroscuro” technique
- System responds with textual explanation of the technique and a number of example images of its application in Rembrandt’s paintings

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**

Multimedia differs from text

- Different document and presentation abstractions
  - hard to separate style from structure
- Formatting is not based on text flow
  - no pages or scrollbars, no line-breaking or hyphenation
  - templates often do not work well either
- Feedback from the formatting back-end required
  - need to check whether proposed layout is feasible
  - layout of media items is less flexible than text layout
- Transformations are hard in a functional language
  - need to try out designs and backtrack when necessary

Cuypers multimedia generation engine

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

Cuypers – the bad news

Currently all our design knowledge is:

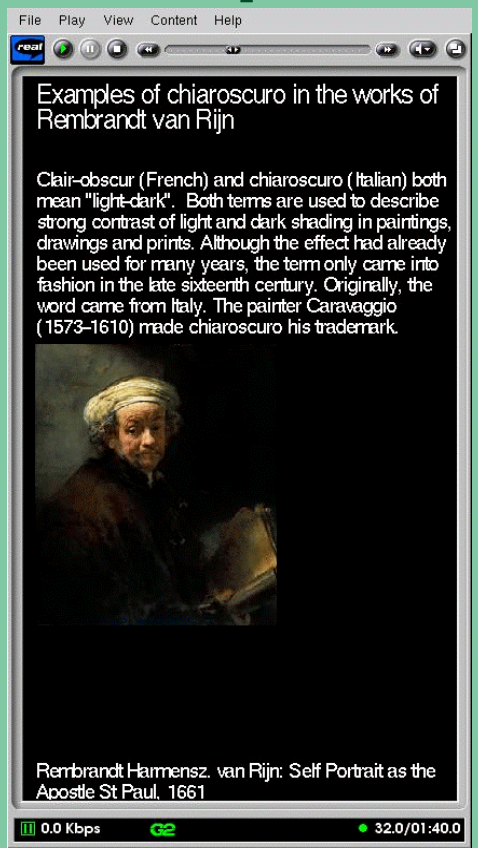
- implicit and hidden in the generation rules
- lost in the generated Web presentation
- not reusable for other Web applications/sites

We need the Semantic Web

Talk overview

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

Example - Annotated SMIL



Using an existing ontology

See http://www.cwi.nl/~media/semantics/

```
<?xml version="1.0"?>
<!-- taken from
   http://www.ics.forth.gr/proj/isst/RDF/RQL/rql.html
   -->
<rdf:RDF xml:lang="en"
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:rdfs="http://www.w3.org/TR/2000/CR-rdf-schema-20000327#"
   xmlns="">

    <rdfs:Class rdf:ID="Artist"/>
    <rdfs:Class rdf:ID="Artifact"/>
    <rdfs:Class rdf:ID="Museum"/>
    <rdfs:Class rdf:ID="Painter">
       <rdfs:subClassOf rdf:resource="#Artist"/>
    </rdfs:Class>
    <rdfs:Class rdf:ID="Painting">
       <rdfs:subClassOf rdf:resource="#Artifact"/>
    </rdfs:Class>
...
</rdf:RDF>
```

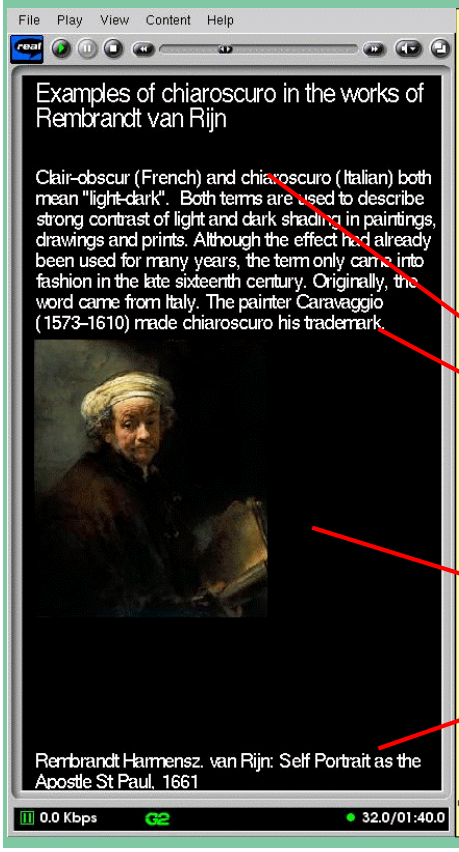
Embedding RDF in SMIL - I

```
<smil xmlns="http://www.w3.org/2000/SMIL20/CR">
  <head>
    <meta name="generator" content="CWI/Cuypers 1.0"/>
    <metadata>
      <rdf:RDF xml:lang="en"
               xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
               xmlns:oil="http://www.ontoknowledge.org/oil/…"
               xmlns:museum="http://ics.forth.gr/.../museum.rdf"

          <museum:Museum rdf:ID="Rijksmuseum" />

          <museum:Painter rdf:ID="Rembrandt">
            <museum:fname>Rembrandt</museum:fname>
            <museum:lname>Harmenszoon van Rijn</museum:lname>
            <museum:paints rdf:resource="#apostlePaul" />
          </museum:Painter>

          <museum:Painting rdf:about="#apostlePaul">
            <museum:exhibited rdf:resource="#Rijksmuseum" />
            <museum:technique>chiaroscuro</museum:technique>
          </museum:Painting>
        </rdf:RDF>
    </metadata>
    . . .
```

Embedding RDF in SMIL - II

```
<museum:Painting rdf:about="#apostlePaul">
          <museum:exhibited rdf:resource="#Rijksmuseum" />
          <museum:technique>chiaroscuro</museum:technique>
          <token:painted-by rdf:resource="#Rembrandt" />
        </museum:Painting>
      </rdf:RDF>
    </metadata>
...
  </head>
  <body>
    <par>
      <text region="title" src="...query to MM DBMS..."/>
      <text region="descr" src="..."/>
      <seq>
        <par dur="10"> ... 1st painting+title ... </par>
        <par dur="10"> ... 2nd painting+title ... </par>
        <par dur="10"> ... 3rd painting+title ... </par>
        <par dur="10"> ... 4th painting+title ... </par>
        <par dur="10" id="apostlePaul">
          <img  region="img" src="..."/>
          <text region="ptitle" src=".."/>
        </par>
      </seq>
    </par>
  </body>
</smil>
```

Marked-up presentation




```
   <museum:Painting rdf:about="#apostlePaul">
        <museum:exhibited rdf:resource="#Rijksmuseum"/>
        <museum:technique>chiaroscuro</museum:technique
   >
   </museum:Painting>
   </rdf:RDF>
   </metadata>
...
  </head>
  <body>
    <par>
      <text region="title" src="..."/>
      <text region="descr" src="..."/>
      <seq>
        <par dur="10"> ... </par>
        <par dur="10"> ... </par>
        <par dur="10"> ... </par>
        <par dur="10"> ... </par>
        <par dur="10" id="apostlePaul">
          <img  region="img" src="..."/>
          <text region="ptitle" src=".."/>
        </par>
      </seq>
    </par>
  </body>
</smil>
```