Multimedia Documents on the Semantic Web

Jacco van Ossenbruggen

CWI

Multimedia and Human-Computer Interaction
http://media.cwi.nl/

Lynda Hardman, Frank Nack & Lloyd Rutledge,
Joost Geurts, Stefano Bocconi

Agenda

1. Research overview
   - Where we come from
   - What we do with multimedia
   - Why we need semantics
2. Practical issues
   - Media/metadata links
   - Multimedia annotation vocabularies
   - Tool and production chain integration

What was the Web about?

A standard means of

- locating documents:
  - URL (URI,URN,IRI,…)
- transferring documents:
  - HTTP (FTP,Gopher,WAIS,RTSP,…)
- encoding documents:
  - HTML (XML,XHTML,SVG,SMIL)

3 generations web documents

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
   - explicit meta-data instead of "screen scraping"
   - agreed upon semantics (RDFS, OWL, DC, …)

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!**

We need 2nd generation multimedia

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

Slide 7/46

Why multimedia is different from text…

- Different document and presentation abstractions
  - hard to separate style from structure & content
- 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-based layout
- Transformations are hard in a functional language
  - need to try out designs and backtrack when necessary

Slide 8/46

But multimedia is different from text…
Different presentation abstractions

- hard to separate style from structure & content
- 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-based layout
- Transformations are hard in a functional language
  - need to try out designs and backtrack when necessary

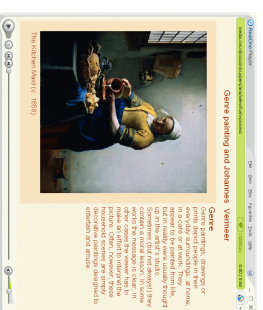# Space/time trade-offs in Cuypers

---

Space/time trade-offs in Cuypers

- Media repository from Rijksmuseum
- Quantitative constraints necessary for pixel-based positioning & synchronization, but insufficient
- Qualitative constraints also used for specification of constraints at higher level
  `A not-overlap B, B after C`
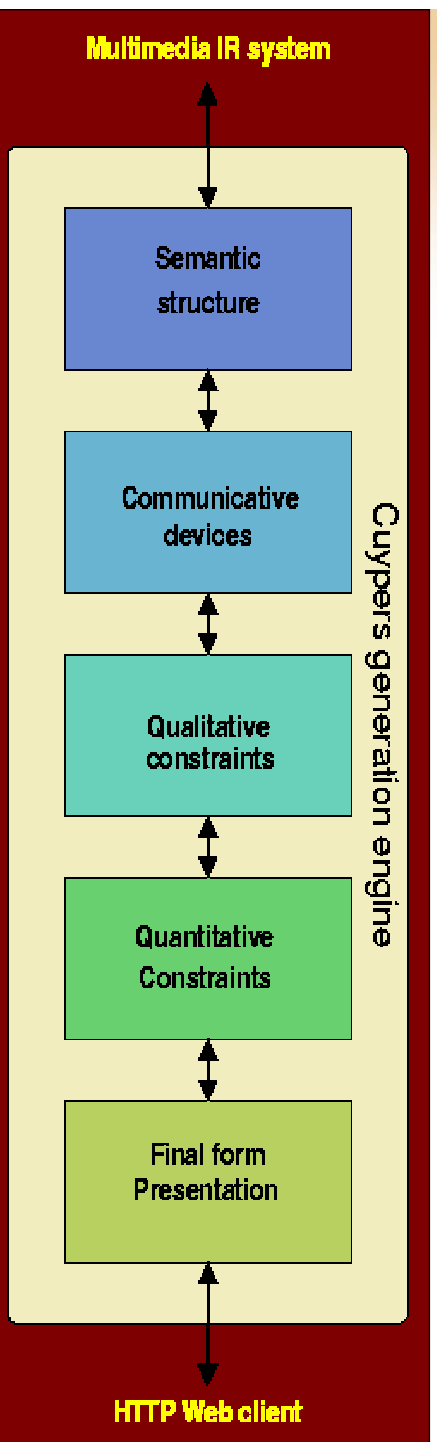- If insoluble then backtrack to other solutions using Prolog
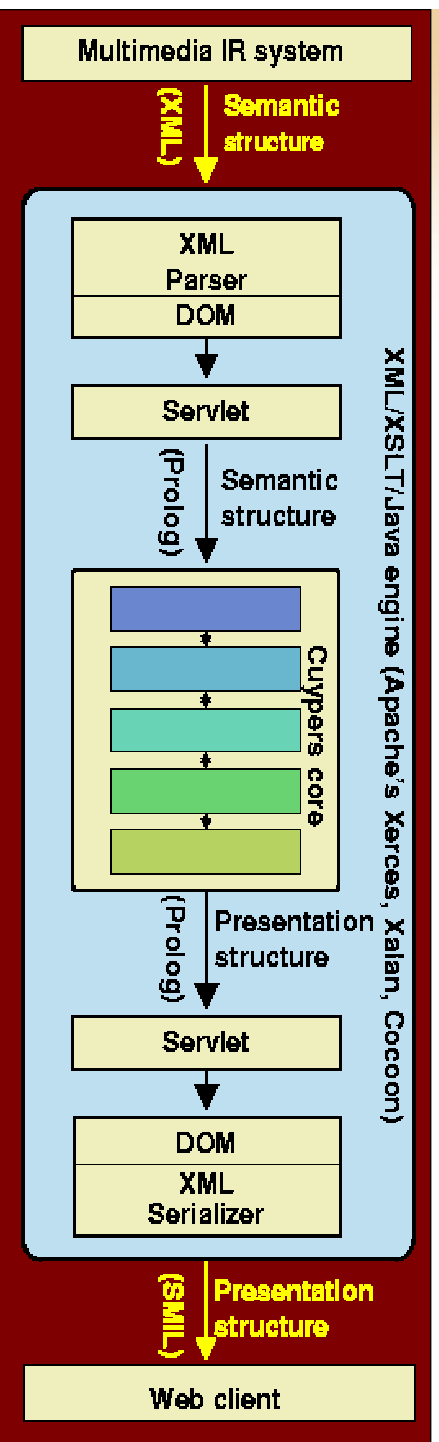
- Geurts & van Ossenbruggen MMM 2001, WWW 2001

Cuypers Web Embedding

Multimedia IR system

(XML) **Semantic structure**

XML Parser
DOM

↓

Servlet

(Prolog) Semantic structure

↓

Cuypers core

(Prolog) Presentation structure

↓

Servlet

↓

DOM
XML Serializer

XML/XSLT/Java engine (Apache's Xerces, Xalan, Cocoon)

(SMIL) **Presentation structure**

Web client

- Prolog-based engine embedded in Apache
- XML-based input/output stream
- Java servlets for XML to Prolog translation

Cuypers multimedia presentation generation engine

Multimedia IR system

Semantic structure

↕

Communicative devices

↕

Qualitative constraints

↕

Quantitative Constraints

↕

Final form Presentation

Cuypers generation engine

HTTP Web client

- Cuypers is based on
  - media independent presentation abstractions
  - transformation rules with built-in backtracking and constraint solving (CLP)

Document Engineering Perspective

- Content is selected & structured (e.g. XML)
- Mappings are defined to a new presentation structure (e.g. SMIL)
- Styles can be applied (such as color and fontsizes)
- The transformation process is linear and assumes that
  - Content/document structure,
  - presentation structure and,
  - style

  are independent of each other.

But in multimedia, space/time design is about semantics!

- Content provider
  - Mission (make profit, promote image)
  - Limited resources (cost)
  - Preferences (company colors)
- End-user
  - Goals, needs
  - Delivery context limitations (time, environment)
  - Preferences (images vs text, audio vs visual)
- Designer
  - Design experience
  - Resource limitations

"Graphic Design Perspective"

- Basically:
  presentation structure, content and style
  depend on one another.

- In multimedia presentations, spatio-temporal layout gives meaning to the presentation by conveying its structure and the semantic relations among the media items

Presentation
structure

Content

Aesthetics

abcdefghij
abcdefghijk
ABCDEFGHIJK
abcdefghijk
abcdefghijk
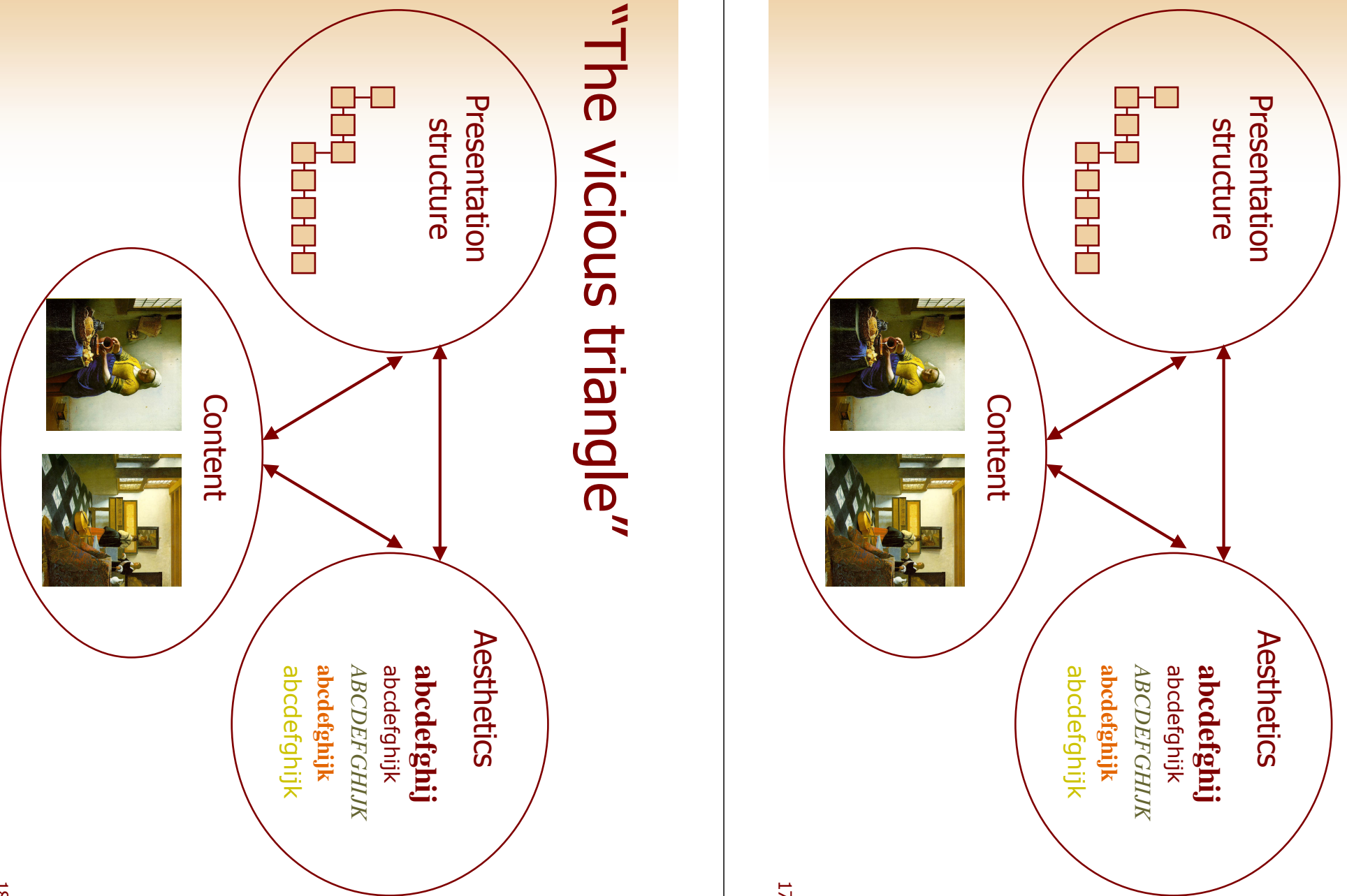
"The vicious triangle"

Presentation
structure

Content

Aesthetics

abcdefghij
abcdefghijk
ABCDEFGHIJK
abcdefghijk
abcdefghijk

Inferring document structure

Rijksmuseum Artifacts Matching the Query " beach "

Pinks in the Breakers

Pinks in the Breakers
by Hendrik Willem Mesdag
c. 1875-1885

Fishing boats are lying at anchor in the surf before Scheveningen. They are 'pinks,' flat-bottomed boats with no keel and a broad curved prow. They were used in the herring fishery. The boats would be dragged along the beach by horses. Scheveningen did not yet have a harbour in the nineteenth century. To the left is a pink on dry land, surrounded by fisherfolk, mainly women with baskets of fish. A strong breeze is blowing and the breakers are capped with white foam. Mesdag was skilled at catching the feeling of a rough day at

Presentation structure
depends on content

genre

Vermeer

1665

1665

1658

cityscapes

1623

1628
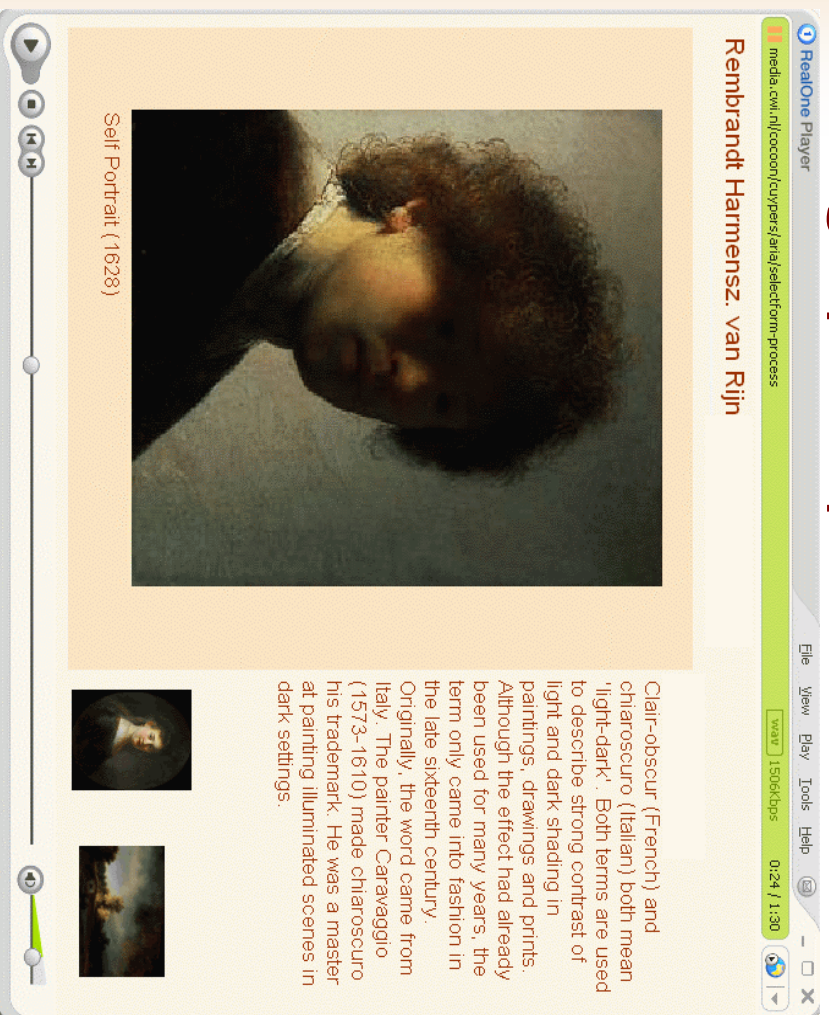
Inferring document structure

- Topia
- Rijksmuseum ARIA database -> RDF
- Clustering on results of query
- Presentation showing "table of contents" and current focus

- Rutledge & Hardman
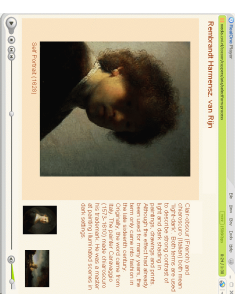 ACM Hypertext 2003
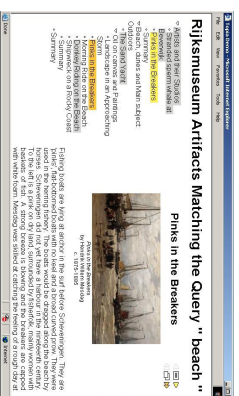
Semantic graph to presentation

Semantic graph to presentation

- DISC
- Rijksmuseum repository of media items
- Semantic graph is not enough
  Rembrandt married-to Saskia
  also need discourse structures
  for deriving grouping, ordering and priorities
- Biography template created
  painter is-a profession

- Stefano Bocconi, Joost Geurts
  ISWC 2003

Argument generation in video

Argument generation in video

- Vox Populi
- Database with "Interview with America" video clips
- RDF-annotated with topic and agree/disagree
- Argumentation model (Toulmin)
- User specifies query and video sequence generated

- Bocconi & Nack,
  ICME 2005, ACM Hypertext 2005

Wrap-up part I:
Scientific challenges

- Making (multimedia) discourse and design knowledge explicit
- Expressing re-usable semantics of media assets
  – Semantic Web
- Designing architectures for multimedia presentation generation

So in all our projects we need the Semantic Web

- Domain & discourse semantics
- – What is this about?
- – Potential functional roles in presentation
  - • intro, conclusion, …
- – Rhetorical relations
  - • Example-of, illustration-of, counter-argument-to, …
- Used to
- – Select media items (IR)
- – Construct presentation structure
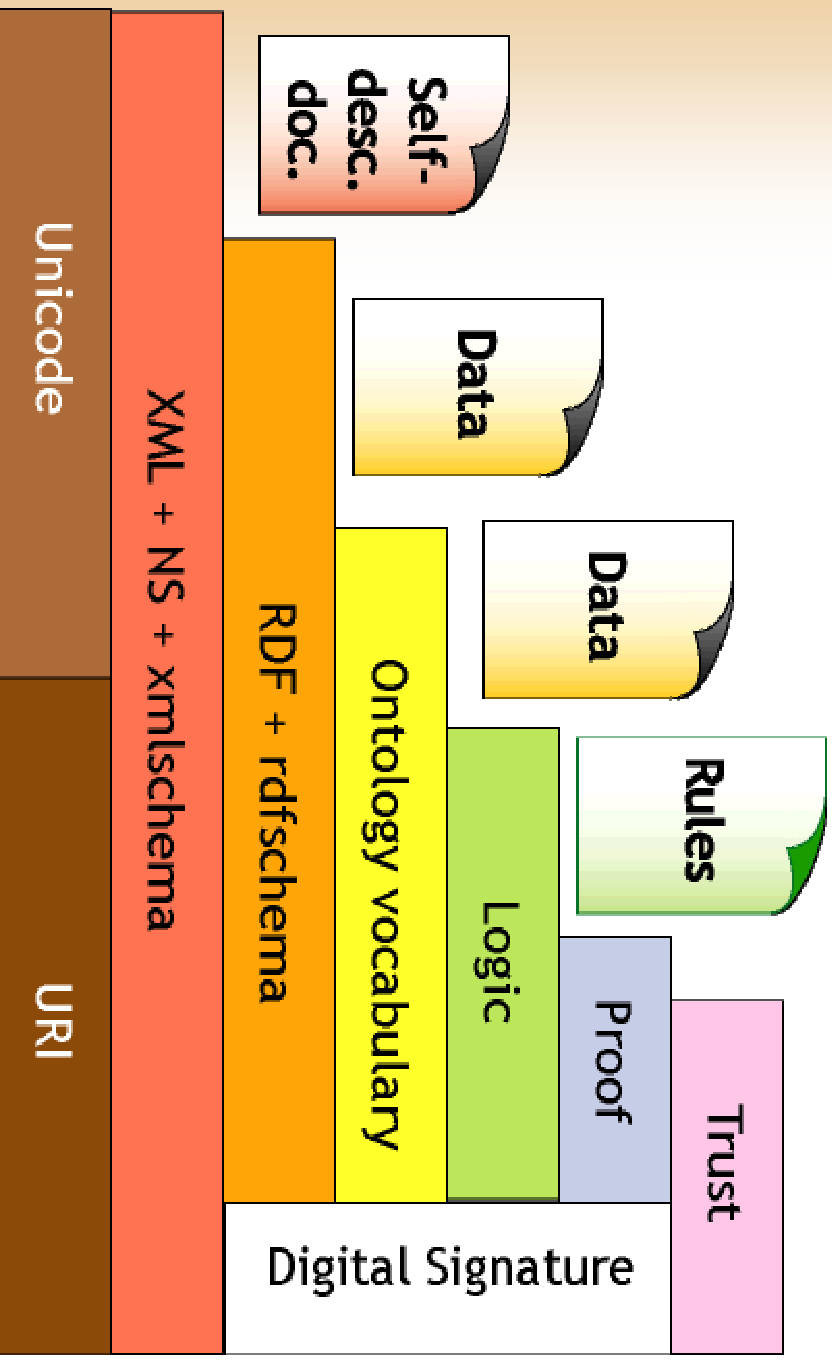- – Make space/time trade-offs
- – Assign style properties

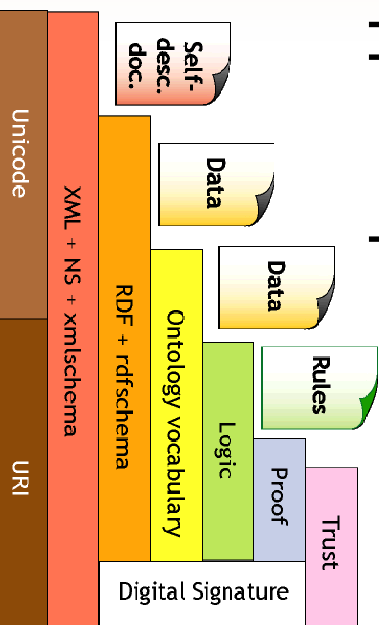Part II

Multimedia and the Semantic Web:
practical issues

Enter TBL's layer cake

Missing bits and pieces

- Binary, streaming non-textual documents?
- Linking data & metadata
- Vocabularies
- Ontological primitives?
- Tool support & production chain integration

Linking data & metadata

- Linking:
  – Attach multiple annotations to single media item (*n:1*)
  – Attach single annotations to multiple media items (*1:m*)
  – (*n:m*)
- Requires flexible linking
  – Beyond *rdf:about* (or: Hypertext linking revisited)
  – External or embedded links (in data, in metadata, or …)
  – Distributed ownership (of data, metadata & links)

Linking data & metadata

- Anchoring: attach metadata to
  – Specific player in football match
  – Audio fragment of gun shot
  – Third character out of shot
  – Specified bounding box in frame smpte=31:21:33:20
  – Video currently playing in active window
  – The last viewed document
- Requires flexible anchoring:
  – Beyond *index.html#foobar*
  – fragment identifiers defined for media types other than *text/xml*…
  – Example: SVG viewports

Linking ("HTML" approach)

`<a name="seq1">`

`<a name="seq2">`

m:clip1.mpg#seq1 - "campanile"
m:clip1.mpg#seq2 - "close up of campanile."

Linking (current, only rdf:about)

http://.../clip1.mpg (m:clip1.mpg)

document

annotation
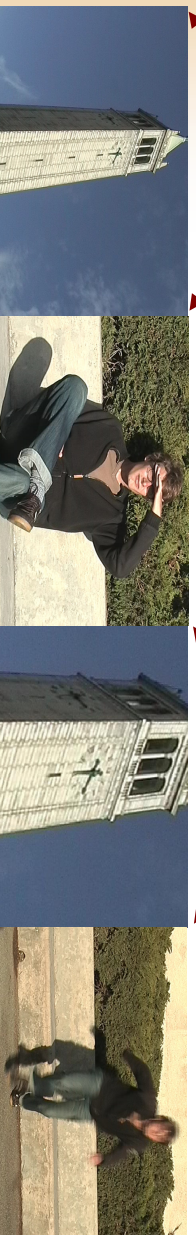
m:clip1.mpeg - "Joost sitting on a bench looking at the campanile clock tower. Then jumps up as he realizes he is late"
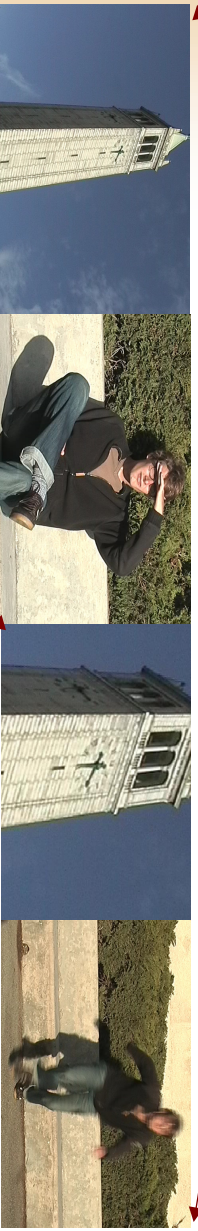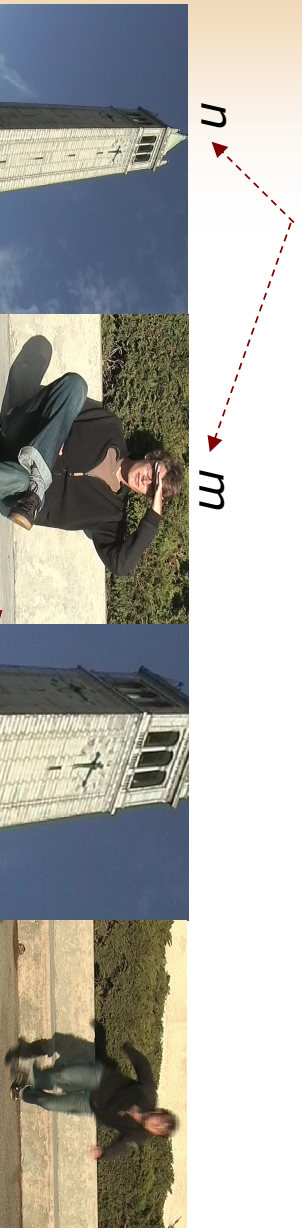
Linking (MPEG7 approach)

http://.../clip1.mpg

Id – myclip
myclip.URL – m:clip1.mpg
myclip.start - 0
myclip.end – 50
myclip.description - "Campanile in Berkeley"

Id – mycl
myclip.U
myclip.st
myclip.en
myclip.de

Linking (URI approach)

http://.../clip1.mpg#n-m

n

m

http://.../clip1.mpg#0-90 – "Establishing shot"

http://.../clip1.mpg#70-150 – "Reaction shot"

Vocabularies: Aplication scenario: Paintings

Slide 37

Knowledge corpora
AAT
ULAN
ICONCLASS
WordNet

Annotation Template
VRA 3.0
Scene descriptors
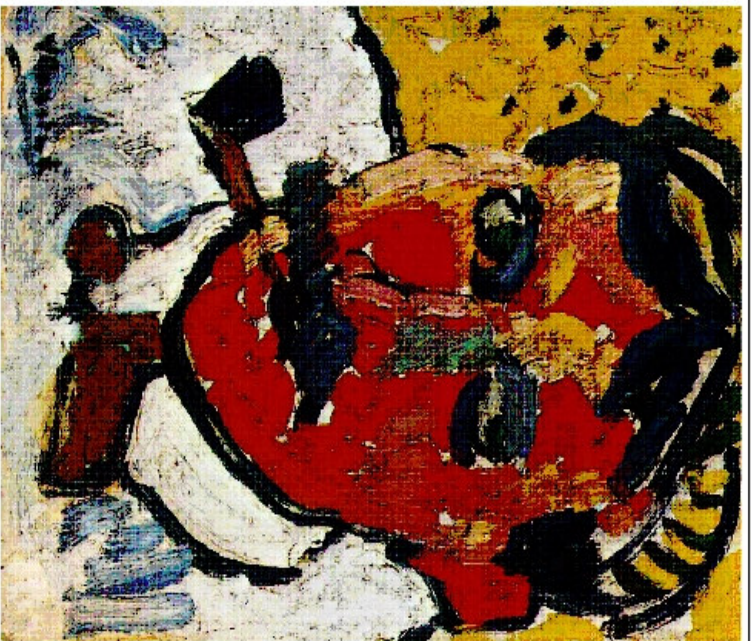
RDF Schema

Annotation & search tool

RDF image annotations

This and next two slides courtesy of Guus Schreiber, VU Amsterdam

Slide 38

MIA

File   Annotate   Search   Ontology   Help

Found "pipe" 9 times in ontology

Production | Physical | Content | Administration

Create: | Active Agent | Objects | Setting

Information_Object

Active Agent

Agent
Derain, André
With

Action
smoke
With

Object
pipes (smoking equipment)
With

Recipient
With

Ok   Reset

Vocabularies: term disambiguation
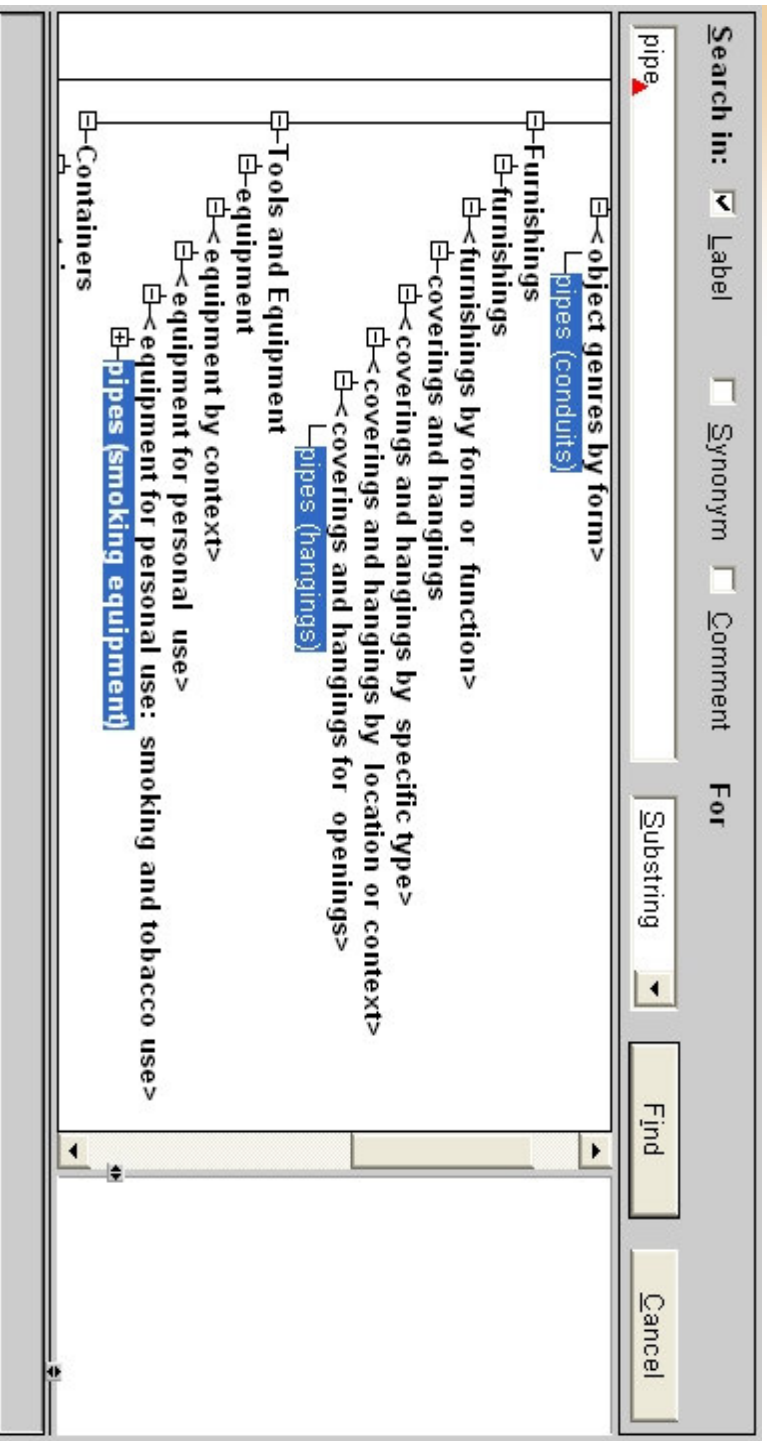
Vocabularies: observations

- Multimedia applications typically use
  - not one "multimedia ontology"
  - but multiple semantic sources: thesauri, vocabularies, ontologies
- Semantic web languages solve the syntactic interoperability problem
- What remains is linking the semantics!
  - annotation templates

Vocabularies: misconceptions

- Ontology instantiation ≠ media annotation
- RDF/OWL schema ≠ annotation template
- Ontology editor ≠ annotation tool
- All metadata ≠ RDF
  - Leverage current practice: ID3, EXIF, DC, VRA
  - MPEG7
- A single vocabulary is insufficient
  - Combining vocabularies is not trivial
  - Namespaces only solve syntax-level name clashes but provide no semantic integration

Fit into current practice

- Metadata support in current tools & production chain
- Keep digital production artefacts and make them available as metadata
  - Storyboards, scripts, edit decision lists
  - "The making of …", DVD bonus material, etc
  - MPEG7 camera of Nack et al.

- Uncertain formation
  - Fuzzy DLs (Stoilos et al.)
- Coordinate spaces
  - space, time, colour, …
  - don't say the H-word ("HyTime" ☺)
- Common multimedia properties
  - sample rates, resolutions, screen sizes, lens positions, etc
  - Align RDF vocabulary with MPEG7 (XSD, Hunter 2001)
- Media-specific
  - e.g. shot/scene/sequence for video
- Media-(record)-metadata relations
  - foaf:depicts, vra:type=work, vra:type=image

Do we really need new technology & more standards?

- Or just need to learn to apply and combine what's already out there
  - Semantic Web stack
  - MPEG stack
  - Other metadata formats
- Put some simple, open, easy to use multimedia vocabularies on the web
- Collect and publish "Best practices" for multimedia on the Semantic Web

Thanks

- Questions?

This research is supported by

- BSIK/MultimediaN

- NWO ToKeN/I²RP
  Intelligent Information Retrieval and Presentation

  ToKeN2000

- NWO ToKeN/CHIME
  Cultural Heritage in an Interactive Multimedia Environment

  ToKeN2000

- NWO NASH
  Networked Adaptive Structured Hypermedia

- Telematica Instituut Topia

- Images courtesy of Rijksmuseum, Amsterdam