A Semantic Multimedia Web: 
Create, Annotate, Present and Share your Media

Lynda Hardman, Raphaël Troncy

<Lynda.Hardman|Raphael.Troncy@cwi.nl>

CWI, Semantic Media Interfaces
Motivation
Learning Objectives

• Understand multimedia applications workflow
  – Take the canonical processes of media production model

• Explore various multimedia metadata formats
  – Be aware of the advantages and limitations of various models
  – Know the interoperability issues and understand COMM, a Core Ontology for Multimedia

• Discuss exploratory interfaces based on rich multimedia metadata semantics
  – Know how to link and expose your data on the web
  – See various multimedia presentation interfaces
Agenda

1. Understanding Multimedia Applications Workflow
   – CeWe Color Photo Book creation application
   – Vox Populi argumentation-based video sequences generation
   – *Canonical Processes of Media Production*

2. Semantic Annotation of Multimedia Content
   – Multimedia metadata formats: use cases and requirements
   – Multimedia metadata interoperability issues
   – MPEG-7 based ontologies
   – *COMM: A Core Ontology for MultiMedia*

3. Semantic Search and Presentation of Multimedia Content
   – Link your data!
   – *Searching and Browsing Multimedia Semantic Datasets with Cliopatria*
Understanding Multimedia Applications Workflow

• Identify and define a number of canonical processes of media production

• Community effort
  – 2005: Dagstuhl seminar
  – 2008: Multimedia Systems Journal Special Issue (core model and companion system papers)
    editors: Frank Nack, Zeljko Obrenovic and Lynda Hardman
Overview of Canonical Processes

- Premeditate
- Create
- Annotate
  - Package
  - Construct Message
  - Query
- Organize
- Publish
- Distribute
Example 1: CeWe Color PhotoBook

• Application for authoring digital photo books
• Automatic selection, sorting and ordering of photos
  – Context analysis methods: timestamp, annotation, etc.
  – Content analysis methods: color histograms, edge detection, etc.
• Customized layout and background
• Print by the European leader photo finisher company

http://www.cewe-photobook.com
CeWe Color PhotoBook Processes

- My winter ski holidays with my friends

Premeditate

Construct Message
CeWe Color PhotoBook Processes

- Select your images
- Select book type
- Create your photobook in just a few steps!
CeWe Color PhotoBook Processes

- Annotate
CeWe Color PhotoBook Processes

- Query
  - Use the following characteristics when selecting your images:
    - Avoid similar images
    - Avoid blurred and unsharp images
    - Avoid low resolution images
    - Favour images with bright colours
    - Take personal rating (from Windows) into account.

- Organize
  - Automatic image selection
  - Use the following types of layout:
    - Use largest images possible per page (fewer background)
    - Random distribution of images
    - Tilted or overlapping images
    - Panorama images

CeWe Color PhotoBook Processes

- Publish
- Distribute
CeWe Color PhotoBook Processes
Example 2: Vox Populi Video Sequences Generation

Stefano Bocconi, Frank Nack

• **Interview with America**
  video footage with interviews and background material about the opinion of American people after 9-11
  [http://www.interviewwithamerica.com](http://www.interviewwithamerica.com)

• Example question:
  *What do you think of the war in Afghanistan?*

  “I am never a fan of military action, in the big picture I don’t think it is ever a good thing, but I think there are circumstances in which I certainly can’t think of a more effective way to counter this sort of thing...”
Vox Populi Premeditate Process

- Analogous to the pre-production process in the film industry
  - Static versus dynamic video artifact
- Output
  - Script, planning of the videos to be captured
  - Questions to the interviewee prepared
  - Profiles of the people interviewed: education, age, gender, race
  - Locations where the interviews take place
Vox Populi Annotations

• Contextual
  – Interviewee (social), locations

• Descriptive
  – Question asked and transcription of the answers
  – Filmic continuity, examples:
    • gaze direction of speaker (left, centre, right)
    • framing (close-up, medium shot, long shot)

• Rhetorical
  – Rhetorical Statement
  – Argumentation model: Toulmin model
Vox Populi Statement Annotations

- Statement formally annotated:
  - \(<subject> <modifier> <predicate>\)
  - E.g. “war best solution”

- A thesaurus containing:
  - Terms on the topics discussed (155)
  - Relations between terms: similar (72), opposite (108), generalization (10), specialization (10)
  - E.g. war opposite diplomacy
Toulmin Model

57 Claims, 16 Data, 4 Concessions, 3 Warrants, 1 Condition
### Vox Populi Query Interface

<table>
<thead>
<tr>
<th>Question</th>
<th>Interviewee</th>
<th>Opinion</th>
</tr>
</thead>
<tbody>
<tr>
<td>What do you think of the Afghanistan war?</td>
<td>Cameroun Parking Guard at Stamford</td>
<td>War in Afghanistan - Pro</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>Education</th>
<th>Employment</th>
<th>GeoLocation</th>
<th>Race</th>
<th>Religion</th>
<th>Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>MiddleAge</td>
<td>Old</td>
<td>Retired Student</td>
<td>NotUSA</td>
<td>AmericanIndian</td>
<td>Atheist</td>
<td>Female</td>
</tr>
<tr>
<td>Old</td>
<td>LowEducated</td>
<td>MiddelIncomeJob</td>
<td>USA</td>
<td>Asian</td>
<td>Christian</td>
<td>Male</td>
</tr>
<tr>
<td>Teenager</td>
<td>HighEducated</td>
<td>MiddelIncomeJob</td>
<td>USA</td>
<td>Black</td>
<td>Muslim</td>
<td>Male</td>
</tr>
<tr>
<td>Young</td>
<td>MediumEducated</td>
<td>Retired Student</td>
<td>USA</td>
<td>Hispanic</td>
<td>Unknown</td>
<td>Male</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Bandwidth</th>
<th>Intercut</th>
<th>Caption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create Clash</td>
<td>Low Bandwidth</td>
<td>True</td>
<td>On (can cause problems)</td>
</tr>
<tr>
<td>Create Support</td>
<td>Medium Bandwidth</td>
<td>False</td>
<td>Off</td>
</tr>
<tr>
<td>Vox Populi</td>
<td>High Bandwidth</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Construct Message**

**Position**

**First Character**

**Second Character**

**Query**
Vox Populi Organize Process

- Using the thesaurus, create a graph of related statements
  - nodes are the statements (corresponding to video segments)
    - “war best solution”,
    - “diplomacy best solution”,
    - “war not solution”
  - edges are either support or contradict
I am not a fan of military actions

I cannot think of a more effective solution

War has never solved anything

Two billions dollar bombs on tents
Vox Populi Processes
Canonical Processes 101

• Canonical: reduced to the simplest and most significant form possible without loss of generality

• Each process
  – short description
  – illustrated with use cases
  – input(s), actor(s) & output(s)

• Formalization of processes in UML diagrams in paper (see literature list)
Premeditate

• Establish initial ideas about media production
  – *Design a photo book of my last holidays for my family*
  – *Create argument-based sequences of videos of interviews after September 11*

• Inputs: ideas, inspirations from human experience

• Actors:
  – camera owner
  – group of friends

• Outputs:
  – decision to take camera onto ski-slope
  – structured set of questions and locations for interviews
Create Media Asset

• Media assets are captured, generated or transformed
  – Photos taken at unspecified moments at holiday locations
  – Synchronized audio video of interviewees responding to fixed questions at many locations

• Inputs:
  – decision to take camera onto ski-slope;
  – structured set of questions and locations for interviews

• Actors:
  – (video) camera, editing suite

• Outputs:
  – images, videos
Annotate

• Annotation is associated with asset
• Inputs:
  – photo, video, existing annotation
  – optional thesaurus of terms
• Actors:
  – human, feature analysis program
• Outputs:
  – Complex structure associating annotations with images, videos

Q: “What do you think of the Afghanistan war?”
Speaker: Female, Caucasian...
Semantic Annotate

- Annotation uses existing controlled vocabularies
  - *Subject matter annotations of your photos (COMM, XMP)*
  - *Rhetorical annotations in Vox Populi*

![Diagram showing subject, modifier, and predicate thesauri with war, best, and solution terms]
Package

- Process artifacts are packed logically or physically
- Useful for storing collections of media after capturing...
- ... before selecting subset for further stages
Query

- User retrieves a set of process artifacts based on a user-specified query
- Inputs:
  - user query, in terms of annotations or by example
  - collection(s) of assets
- Actors:
  - human
- Output:
  - subset of assets plus annotations (in no order)
Construct Message

• Author specifies the message they wish to convey
  – *Our holiday was sporty, great weather and fun*
  – *Create clash about whether war is a good thing*

• Inputs: ideas, decisions, available assets

• Actors:
  – author

• Outputs:
  – the message that should be conveyed by the assets
Organize

• Process where process artifacts are organized according to the message
  – Organize a number of 2-page layouts in photobook
  – Use semantic graph to select related video clips to form linear presentation of parts of argument structure

• Inputs: set of assets and annotations (e.g. output from query process)
• Actors: human or machine
• Outputs: document structure with recommended groupings and orderings for assets
Publish

- Presentation is created
  - associated annotations may be removed
    - create proprietary format of photobook for upload
    - create SMIL file containing videos and timing information

- Inputs: set of assets and annotations
  (e.g. output from organize process)

- Actors: human or machine

- Outputs:
  - final presentation in specific document format, such as html, smil or pdf
Distribute

- Presentation is transported to end user, end-user can view and interact with it
  - photobook uploaded to printer, printed then posted to user
  - SMIL file is downloaded to client and played

- Inputs: published document (output from publish process)
- Actors: distribution hardware and software
- Outputs:
  - media assets presented on user’s device
Canonical Processes Possible Flow
Sum Up

- Community agreement, not “yet another model”
- Large proportion of the functionality provided by multimedia applications can be described in terms of this model
- Initial step towards the definition of open web-based data structures for describing and sharing semantically annotated media assets
Discussion

• Frequently asked questions
  – Complex processes
  – Interaction
  – Complex artifacts and annotations can be annotated

• Towards a more rigorous formalization of model
  – Relationship to foundational ontologies
  – Semantics of Annotations
Literature


- Special Issue on Canonical Processes of Media Production
  [http://www.springerlink.com/content/j0l4q337581652t1/](http://www.springerlink.com/content/j0l4q337581652t1/)
  [http://www.cwi.nl/~media/projects/canonical/](http://www.cwi.nl/~media/projects/canonical/)


Agenda

1. Understanding Multimedia Applications Workflow
   - CeWe Color Photo Book creation application
   - Vox Populi argumentative video sequences generation system
   - *The Canonical Processes of Media Production*

2. Semantic Annotation of Multimedia Content
   - Multimedia metadata formats: use cases and requirements
   - Multimedia metadata interoperability issues
   - MPEG-7 based ontologies
   - *COMM: A Core Ontology for MultiMedia*

3. Semantic Search and Presentation of Multimedia Content
   - Link your data!
   - *Searching and Browsing Multimedia Semantic Datasets with Cliopatria*
The Importance of the Annotations
Multimedia: Description methods

ISO
- MPEG-1
- MPEG-2
- MPEG-4
- MPEG-7
- MPEG-21

W3C
- User Interface & applications
- Proof
- Unifying Logic
- Query: SPARQL
- Ontology: OWL
- Rules: RIF
- RDF-S
- Data interchange: RDF
- XML
- URI

ISO
- MPEG-1
- MPEG-2
- MPEG-4
- MPEG-7
- MPEG-21
MPEG-7: a multimedia description language?

- ISO standard since December of 2001
- Main components:
  - Descriptors (Ds) and Description Schemes (DSs)
  - DDL (XML Schema + extensions)
- Concern all types of media

**Content organization**

**Models**

**Navigation & Access**

**User interaction**

**Creation & Production**

**Usage**

**Media**

**Content management**

**Content description**

**Structural aspects**

**Semantic aspects**

**Part 5 – MDS**

*Multimedia Description Schemes*
MPEG-7 and the Semantic Web

- MDS Upper Layer represented in RDFS
  - 2001: Hunter
  - Later on: link to the ABC upper ontology
- MDS fully represented in OWL-DL
  - 2004: Tsinaraki et al., DS-MIRF model
- MPEG-7 fully represented in OWL-DL
  - 2005: Garcia and Celma, Rhizomik model
  - Fully automatic translation of the whole standard
- MDS and Visual parts represented in OWL-DL
  - 2007: Arndt et al., COMM model
  - Re-engineering MPEG-7 using DOLCE design patterns
Requirements

• MPEG-7 compliance
  – Support most descriptors (decomposition, visual, audio)

• Syntactic and Semantic interoperability
  – Shared and formal semantics represented in a Web language (OWL, RDF/XML, RDFa, etc.)

• Separation of concerns
  – Domain knowledge versus multimedia specific information

• Modularity
  – Enable customization of multimedia ontology

• Extensibility
  – Enable inclusion of further descriptors (non MPEG-7)
# MPEG-7 Based Ontologies

<table>
<thead>
<tr>
<th>Foundational Ontologies</th>
<th>Hunter</th>
<th>DS-MIRF</th>
<th>Rhizomik</th>
<th>COMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td></td>
<td>None</td>
<td>None</td>
<td>DOLCE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Complexity</th>
<th>OWL-Full</th>
<th>OWL-DL</th>
<th>OWL-DL</th>
<th>OWL-DL</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Coverage</th>
<th>MDS+Visual</th>
<th>MDS+CS</th>
<th>All</th>
<th>MDS+Visual</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Applications</th>
<th>Digital Libraries</th>
<th>Digital Libraries</th>
<th>Digital Rights</th>
<th>MM Analysis</th>
</tr>
</thead>
</table>
Common Scenario

The "Big Three" at the Yalta Conference (Wikipedia)
Common Scenario: Tagging Approach

The "**Big Three**" at the Yalta Conference (Wikipedia)

- **Localize a region**
  - Draw a bounding box, a circle around a shape

- **Annotate the content**
  - Interpret the content
  - Tag: Winston Churchill, UK Prime Minister, Allied Forces, WWII
Common Scenario: SW Approach

- Localize a region
  - Draw a bounding box, a circle around a shape
- Annotate the content
  - Interpret the content
  - Link to knowledge on the Web

```
:Reg1 foaf:depicts dbpedia:Winston_Churchill
dbpedia:Winston_Churchill skos:altLabel "Sir Winston Leonard Spencer-Churchill"
dbpedia:Winston_Churchill rdf:type foaf:Person
```

The "Big Three" at the Yalta Conference (Wikipedia)
The Big Three at the Yalta Conference


mpeg7:image

mpeg7:media_locator

mpeg7:spatial_decomposition

Reg1

mpeg7:spatial_mask

mpeg7:dominant_color

rdf:type

rgb(25,255,255)

dbpedia:Churchill

mpeg7:coords

5 25 10 20 15 15 10 10 5 15"^^xsd:string

mpeg7:depicts

mpeg7:still_region
DS-MIRF MPEG-7 Ontology


mpeg7:MediaLocator

mpeg7:image

mpeg7:CreationInformation

mpeg7:Creation

mpeg7:Title

contentString

The Big Three at the Yalta Conference

mpeg7:CreationInformation

mpeg7:SpatialDecomposition

Reg1

mpeg7:SubRegion

mpeg7:Creation

mpeg7:Coords

mpeg7:Polygon

mpeg7:dim

5 25 10 20 15 15 10 10 5 15"^^xsd:string

dbpedia:Churchill

mpeg7:RelatedMaterial

mpeg7:StillRegion

mpeg7:MediaURI


mpeg7:MediaURI

mpeg7:SpatialDecomposition

Reg1

mpeg7:SubRegion

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation

mpeg7:Creation
COMM: Fragment Identification


dns:realized-by

core:image-data
dns:setting

core:semantic-annotation
dns:defines

dns:played-by

do:Person

dns:played-by

loc:region-locator-descriptor
dns:defines

dns:played-by

loc:spatial-mask-role

dns:played-by

loc:bounding-box
data:has-rectangle

5 25 10 20 15 10 10 5 15"^"^xsd:string

dbpedia:Churchill

dns:played-by

rdf:type
Comparison

• Link with domain semantics
  – Hunter: ABC model + \texttt{mpeg7:depicts} relationship
  – DS-MIRF: Domain ontologies needs to subclass the general MPEG-7 categories
  – Rhizomik: Use the \texttt{mpeg7:semantic} relationship
  – COMM: Semantic Annotation pattern

• MPEG-7 coverage
  – Hunter: extension of the MPEG-7 visual descriptors
  – COMM:
    • Formalization of the context of the annotation
    • Representation of the method (algorithm) that provides the annotation
Comparison

• **Modeling Decisions:**
  – DS-MIRF and Rhizomik: 1-to-1 translation from MPEG-7 to OWL/RDF
  – Hunter: Simplification and link to the ABC upper model
  – COMM: NO 1-to-1 translation
    • Need for patterns: use DOLCE, a well designed foundational ontology as a modeling basis

• **Scalability:**

<table>
<thead>
<tr>
<th></th>
<th>Hunter</th>
<th>DS-MIRF</th>
<th>Rhizomik</th>
<th>COMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triples</td>
<td>11</td>
<td>27</td>
<td>20</td>
<td>19</td>
</tr>
</tbody>
</table>
Summary

Semantic descriptions of non-textual media available on the web can be used to facilitate retrieval and presentation of media assets and documents containing them. While technologies for multimedia semantic descriptions already exist, there is as yet no formal description of a high quality multimedia ontology that is compatible with existing (semantic) web technologies. We propose COMM - A Core Ontology for Multimedia based on both the MPEG-7 standard and the DOLCE foundational ontology.

The research is partially supported by the European Commission under contracts:

- FP6-027026, Knowledge Space of semantic inference for automatic annotation and retrieval of multimedia content - K-Space,
- FP6-026978, X-Media Integrated Project.

People

- Thomas Franz
- Steffen Staab
- Raphael Troncy
- Richard Arndt
Scenario: Image

The "Big Three" at the Yalta Conference (Wikipedia)

- Localize a region (bounding box)
- Annotate the content (interpretation)
  - Tag: Winston Churchill, UK Prime Minister, Allied Forces, WWII
  - Link to knowledge on the Web

```
:Reg1 foaf:depicts dbpedia:Winston_Churchill

dbpedia:Winston_Churchill skos:altLabel "Sir Winston Leonard Spencer-Churchill"

dbpedia:Winston_Churchill rdf:type foaf:Person
```
Scenario: Video

- Localize a region
- Annotate the content
  - Tag: G8 Summit, Heiligendamm, 2007
  - Link to knowledge on the Web

```
:Seq1 foaf:depicts dbpedia:34th_G8_Summit
:Seq4 foaf:depicts dbpedia:EU_Summit
go:Heilegendamn skos:broader go:Germany
```

A history of G8 violence (© Reuters)
EU Summit, Gothenburg, 2001
Research Problem

• Multimedia objects are complex
  – Compound information objects, fragment identification
• Semantic annotation
  – Subjective interpretation, context dependent
• Linked data principle
  – Open to reuse existing knowledge

The "Big Three" at the Yalta Conference (Wikipedia)

A history of G8 violence (video)
(© Reuters)
COMM: Design Rationale

• Approach:
  – NO 1-to-1 translation from MPEG-7 to OWL/RDF
  – Need for patterns: use DOLCE, a well designed foundational ontology as a modeling basis

• Design patterns:
  – Ontology of Information Objects (OIO)
    • Formalization of information exchange
    • Multimedia = complex compound information objects
  – Descriptions and Situations (D&S)
    • Formalization of context
    • Multimedia = contextual interpretation (situation)

• Define **multimedia patterns** that translate MPEG-7 in the DOLCE vocabulary
COMM: Core Functionalities

• Most important MPEG-7 functionalities:
  – **Decomposition** of multimedia content into segments
  – **Annotation** of segments with metadata
    • Administrative metadata: creation & production
    • Content-based metadata: audio/visual descriptors
    • Semantic metadata: interface with domain specific ontologies

⇒ Note that all are subjective and context dependent situations
Definition of design patterns for **decomposition** and **annotation** based on D&S and OIO

- MPEG-7 describes digital data (*multimedia information objects*) with digital data (*annotation*)
- *Digital data* entities are information objects
- Decompositions and annotations are *situations* that satisfy the rules of a method or algorithm
COMM: Decomposition Pattern

D&S / OIO

- description
  - structured-data-description
    - descriptor
      - localization-descriptor

- information-object
  - digital-data

- role
  - processing-role
    - output-role
      - segmentation-algorithm
    - input-role

- method
  - algorithm
    - segment-decomposition

- situation

MPEG-7
COMM: Annotation Pattern

![Diagram of MPEG-7 Annotation Pattern]

MPEG-7
COMM: Semantic Pattern
COMM: Modules

Annotation Pattern

Decomposition Pattern
Example 1: Fragment Identification
Example 1: Region Annotation

dns:realized-by

core:image-data
dns:setting

dns:plays
loc:region-locator-descriptor
dns:defines
loc:bounding-box

data:has-rectangle

5 25 10 20 15 15 10 10 5 15"^^xsd:string

dns:played-by
loc:spatial-mask-role
dns:defines

core:semantic-label-role

dns:played-by

foaf:Person

rdf:type
Example 2: Fragment Identification


dns:realized-by

core:image-data

dns:plays

dns:played-by

dns:defines

dns:defines

loc:media-time-point

"1:21"^^xsd:time

data:has-time

loc:media-time-descriptor

loc:temporal-mask-role

loc:media-time-point
Example 2: Sequence Annotation

dns:realized-by
dns:setting
dns:plays
dns:defines
dns:played-by

tgn:Sweden

tgn:Gothenburg

core:image-data
loc:media-time-descriptor
dns:defines
dns:played-by

dns:temporal-mask-role
loc:media-time-point
"1:21"^^xsd:time
data:has-time

core:semantic-label-role
dns:defines
dns:played-by

core:semantic-annotation
dns:defines
dns:played-by

skos:broader
Implementation

- COMM fully formalized in OWL DL
  - Rich axiomatization, consistency check (Fact++v1.1.5)
  - OWL 2.0: qualified cardinality restrictions for number restrictions of MPEG-7 low-level descriptors

- JAVA API available
  - MPEG-7 class interface for the construction of metadata at runtime
KAT Annotation Tool

![Image of KAT Annotation Tool interface]
YouTube Videos Fragments
Making Music Out Of The Social Noise
W3C Multimedia Semantics XG

http://www.w3.org/2005/Incubator/mmsem/
Managing Personal Photos

- Interoperable Image Metadata
  - Combining EXIF, MPEG-7, IPTC and DIG35 metadata using RDF and OWL schemas
Facetting Music Songs

- Interoperable Music and Social Metadata
  - ID3 Tags + low-level features extraction + lastFM recommendations + FOAF profiles + ...
  - Auto-construction of playlist (similar bit rate), Personalization, Browsing music store
Metadata Working Group

SPECIFICATIONS

Guidelines for Handling Image Metadata
Consumer sharing of still images has exploded with the maturing of Internet services for the storage, manipulation, and sharing of pictures. However, the majority of standards related to still images are oriented toward the documentation of the creation of an image or towards professional (e.g. print media) usage and management of images. In addition, the content overlap between the most commonly used standards can result in some confusion. This document describes how best to use existing standards such as Exif, IPTC, and XMP to address the key organizational metadata questions that most consumers have. Download the specification (PDF: 1.7M).
Adobe’s Extensible Metadata Platform (XMP) is a labeling technology that allows you to embed data about a file, known as metadata, into the file itself. More information on how partners and standards are using XMP is available at the XMP website.

**XMP Specifications**

The following specifications are included in the XMP Toolkit zip package. They are available here for convenient reference.

- **Part 1, Data and Serialization Models** (PDF, 375k) covers the basic metadata representation model that is the foundation of the XMP standard format. The Data Model prescribes how XMP metadata can be organized; it is independent of file format or specific usage. The Serialization Model prescribes how the Data Model is represented in XML, specifically RDF.

- **Part 2, Standard Schemas** (PDF, 470k), provides detailed property lists and descriptions for standard XMP metadata schemas; these include general-purpose schemas such as Dublin Core, and special-purpose schemas for Adobe applications such as Photoshop. It also provides information on extending existing schemas and creating new schemas.

- **Part 3, Storage In Files** (PDF, 629k) provides information about how serialized XMP information is stored in files.
Mission

The mission of the Media Annotations Working Group, part of the Video in the Web Activity, is to provide an ontology designed to facilitate cross-community data integration of information related to media objects in the Web, such as video, audio and images.

See also: charter and liaisons to non-W3C groups
Literature

Agenda

1. Understanding Multimedia Applications Workflow
   - CeWe Color Photo Book creation application
   - Vox Populi argumentative video sequences generation system
   - *The Canonical Processes of Media Production*

2. Semantic Annotation of Multimedia Content
   - Multimedia metadata formats: use cases and requirements
   - Multimedia metadata interoperability issues
   - MPEG-7 based ontologies
   - *COMM: A Core Ontology for MultiMedia*

3. Semantic Search and Presentation of Multimedia Content
   - Link your data!
   - *Searching and Browsing Multimedia Semantic Datasets with Cliopatria*
A Giant Graph Open to the World

wp:2006_FIFA_World_Cup#Final

nc:15054000

nar:subject

events:id

nar:location

foaf:depicts

dbpedia:Zidane

geonames:2950159

Linking Open Data Project

- Expose open datasets in RDF
- Set RDF links among the data items for different datasets
- Over 2 billion triples, 3 millions links (March 2008)
- Many many many more ... now

DBpedia

- DBpedia is a community effort to:
  - extract structured "infobox" information from Wikipedia
  - interlink DBpedia with other datasets on the Web
DBpedia

Extracting Infobox Data

http://en.wikipedia.org/wiki/Calgary

$http://dbpedia.org/resource/Calgary$

```
dbpedia:native_name "Calgary"
  dbpedia:altitude "1048"
  dbpedia:population_city "988193"
  dbpedia:populationMetro "1079310"
  mayor_name
    dbpedia:Dave_Bronconnier
  governing_body
    dbpedia:Calgary_City_Council
```

- Altogether 9,100,000 RDF triples
  extracted from 754,000 infoboxes

Christian Bizer et al. DBpedia – Querying Wikipedia Like a Database (May 11, 2007)
Automatic Links Among Open Datasets

Processors can switch automatically from one to the other ...

DBpedia

Geonames

<http://dbpedia.org/resource/Calgary>
  owl:sameAs <http://sws.geonames.org/5913490>;
  ...

<http://sws.geonames.org/5913490>
  owl:sameAs <http://DBpedia.org/resource/Calgary>
  wgs84_pos:lat "51.050112282";
  wgs84_pos:long "-114.085285152";
  sws:population "968460"
  ...

Take Home Message

• Reuse what is there
  – Of course, one could create RDF data manually ... ... but that is unrealistic on a large scale
  – Goal is to generate RDF data automatically when possible and "fill in" by hand only when necessary
    • service to get RDF from flickr images
      http://www.kanzaki.com/works/2005/imgdsc/flickr2rdf
    • service to get RDF from XMP
      http://www.ivan-herman.net/cgi-bin/blosxom.cgi/WorkRelated/SemanticWeb/xmpeextract.html

• Expose what you make
Cultural Heritage Data Cloud
Professional Art Annotation with Thesauri from the Web
This cultural search engine will give you access to artworks from several museum collections. Type a keyword, for example, Derain, calligraphy, or 1867.
Global financial crisis of September–October 2008

The global financial crisis of September–October 2008 is a major ongoing financial crisis, the worst of its kind since the Great Depression. It became prominently visible in September, 2008 with the failure, merger or conservatorships of several large United States-based financial firms. The underlying causes leading to the crisis had been reported in business journals for many months before September, with commentary about the financial stability of leading U.S. and European investment banks, insurance firms and mortgage banks consequent to the subprime mortgage crisis.[1][2]

Beginning with failures of large financial institutions in the United States, it rapidly evolved into a global crisis resulting in a number of European bank failures and declines in various stock indexes, and significant reductions in the market-value of equities (stock), down 27% as of October 24.[3] and commodities worldwide.[4] The crisis has led to a liquidity problem and the de-leveraging of financial institutions especially in the United States and Europe, which further accelerated the liquidity crisis. World political leaders and national ministers of finance and central bank directors have coordinated their efforts to reduce fears but the crisis is ongoing and continues to change. The crisis has roots in the subprime mortgage crisis and is an acute phase of the financial crisis of 2007–2008.

videos
cartoons
Global financial crisis of September–October 2008

From Wikipedia

The global fin 2008 with the many months subprime mort. Beginning with indexes, and s de-leveraging central bank d phase of the fi

Blogs about: Financial Crisis

Brown: "Welfare reform will be intensified"
News Workflow Interoperability

- No integration of media (stories, photo, animation, video)
- Little (or no) context in the news presentation
- Lack of interoperability in the current workflow
Metadata is Key

• (Ultimate) Goal:
  – Provide an environment for searching and browsing contextualized multimedia news information

• Required integration:
  – Data: various media, different forms, various sources
  – Metadata: schema integration, semantic models

• Influence and implications of UI:
  – How to represent semantic multimedia metadata to facilitate presenting information?
  – *in other words* ... What constraints do end-user interfaces put on the modeling of the metadata?
News and Multimedia Formats

NewsML
EventsML
SportsML
News Architecture (NAR)

XMP
BBC
EXIF.org

IPTC
IFRA
CableLabs
COMM
MPEG7

Porting Schemas and Thesauri to the Semantic Web

- Methodologies and tools for building ontologies: ... from scratch
- SKOSification of thesauri in the CH domain:
  - preparation, syntactic and semantic conversion, standardization

⇒ Lack of best practices for modeling ontologies from UML diagrams, integrating ontologies with various thesauri, while taking the end-user interface into account
Building a Semantic Web Infrastructure for News

1. Modeling the NAR ontology
2. Linking with media ontologies
3. Building SKOS thesauri
4. Enriching the metadata
Step 1: Modeling the NAR Ontology

- AnyItem
  - NewsItem
  - PackageItem
  - ConceptItem
  - KnowledgeItem

- Text
- Photo
- Graphics
- Animation
- Audio
- Video
- Composite

- Text
- Photo
- Person Organisation
- Composite

- Person Organisation
- Geopolitical Area
- Point of Interest
- Event
- Composite

⇒ focus on reuse of XML types leading to multiple repetition resulting in overly complex nested XML structures
Step 1: Modeling the NAR Ontology

- Flattening the XML structure

```xml
<!--AFP NewsML2 text-photo profile-->
<!--Processed by Xafp1-4ToNewsML2 b11-->
<newsMessage>
  + <header/></header>
  - <ItemSet>
    - <newsItem guid="urn:newsml:afp.com:20010101:.DV90996" schema="0.7" version="11" xml:lang="en">
      <catalogRef href="http://iptc.org/std-dev/NAR/1.0/specification/IPTC-TempCatalog-inc_3.xml"/>
      - <itemMeta>
        <contentClass code="ccls:photo"/>
        <provider literal="afp.com"/>
        <itemCreated>2006-07-09T21:20:00Z</itemCreated>
        <modified>2006-07-11T09:14:38Z</modified>
        <fileName>DV90996</fileName>
        <edNote>MOBILE SERVICES OUT</edNote>
      </itemMeta>
      + <contentMeta></contentMeta>
      + <contentType></contentType>
    </newsItem>
  </ItemSet>
</newsMessage>
```
Step 1: Modeling the NAR Ontology

• Modeling unique identifiers
  – Use of dereferencable URIs for any resources (news items + vocabularies)
  – Future: Use of URIs for resource fragments
    http://www.youtube.com/watch?v=1bibCui3lFM#t=1m45s

• Modeling the provenance of the information
  – Reification
  – Named (and Networked) Graphs

{<> nar:subject cat:11002000}
  dc:creator team:md ;
  dc:modified ‘‘2005-11-11T08:00:00Z’’.
Step 2: Linking with Media Ontologies

- foaf:Person ≈ nar:Person
- dc:Subject ≈ nar:Subject
- sioc:Item ≈ nar:Item
- geo:lat
- geo:long
Step 3: Getting SKOS Vocabularies
Step 3: Getting SKOS Vocabularies

xml version="1.0" encoding="UTF-8">
<!DOCTYPE rdf:RDF []>
  <skos:Concept rdf:about="http://newsml.cwi.nl/NewsCodes/topicset.iptc-subjectcode/04010010"/>
  <skos:inScheme rdf:resource="http://newsml.cwi.nl/NewsCodes/topicset.iptc-subjectcode/"/>
  <skos:altLabel>04010010</skos:altLabel>
  <skos:prefLabel xml:lang="it">Industria televisiva</skos:prefLabel>
  <skos:definition xml:lang="it">La televisione come industria</skos:definition>
  <skos:changeNote xml:lang="it">nessuno</skos:changeNote>
  <skos:prefLabel xml:lang="fr">Industrie de la télévision</skos:prefLabel>
  <skos:definition xml:lang="fr">Les histoires liées à l'industrie de la télévision</skos:definition>
  <skos:changeNote xml:lang="fr">none</skos:changeNote>
  <skos:prefLabel xml:lang="es">televisión</skos:prefLabel>
  <skos:definition xml:lang="es">Historias relacionadas al negocio de la televisión</skos:definition>
  <skos:changeNote xml:lang="es">none</skos:changeNote>
  <skos:prefLabel xml:lang="de">Fernsehproduktion</skos:prefLabel>
  <skos:definition xml:lang="de">Artikel, die sich (im wirtschaftlichen Sinn) auf das Fernsehen beziehen</skos:definition>
  <skos:changeNote xml:lang="de">none</skos:changeNote>
  <skos:prefLabel xml:lang="en-GB">television industry</skos:prefLabel>
  <skos:definition xml:lang="en-GB">Stories related to the business of television</skos:definition>
  <skos:changeNote xml:lang="en-GB">none</skos:changeNote>
</skos:Concept>
</rdf:RDF>
Step 4: Enriching the News Metadata

- Concepts/Entities that are subject of news
  - Thematic categories
  - People
  - Organizations
  - Geopolitical Areas
  - Points of Interest
  - Events
  - Products or artefacts
Step 4: Enriching the News Metadata
Step 4: Enriching the News Metadata

- Domain Ontologies
- Concept Detectors

NAR Ontology
NewsCodes
Thesaurus

K-Space

Concept Detectors

Domain Ontologies

RDF

WordNet
Web of Data and Linked Data

nc:15054000

nar:subject

events:id

nar:location

foaf:depicts

geonames:2950159
dbpedia:Zidane

wp:2006_FIFA_World_Cup#Final
Presenting News Information

- Dimensions used for searching news items
  - When \textit{time} 10/07/2006
  - Where \textit{location} Paris
  - What \textit{is depicted} J. Chirac, Z. Zidane
  - Why \textit{event} WC 2006
  - Who \textit{photographer} Bertrand Guay, AFP
<table>
<thead>
<tr>
<th>Description</th>
<th>Number of RDF Triples</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Ontologies: NAR, DC, FOAF</td>
<td>7,336</td>
</tr>
<tr>
<td>Domain Specific Ontologies: football</td>
<td>104,358</td>
</tr>
<tr>
<td>Thesauri: newscodes</td>
<td>34,903</td>
</tr>
<tr>
<td>DBpedia, Geonames</td>
<td>53,468</td>
</tr>
<tr>
<td>AFP News Feed (June/July 2006)</td>
<td>804,446</td>
</tr>
<tr>
<td>AFP Photos (June/July 2006)</td>
<td>61,311</td>
</tr>
<tr>
<td>INA Broadcast Video (June/July 2006)</td>
<td>1,932</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1,067,754</strong></td>
</tr>
</tbody>
</table>
This news search engine will give you access to news items kindly provided by AFP.
Type a keyword, for example: Amsterdam, Lyon, Beyrouth, Zidane or G8.

Collections

- AFP.com (> 100,000 objects)
- INA.fr (> 500 objects)

Vocabularies & Thesauri

- IPTC
- DBpedia
- GeoNames
- SWAP
Provide new dimensions for browsing

Distinguish field images vs stadium and street images with a grass detector for the World Cup dataset
### property | value
--- | ---
SecondOrMoreBroadcast | yes
audiencePresent | yes
broadcastGoal | 
broadcastInformation | 13/07/2006 - type: Diffusé - heure: 20:16:01 - canal: 2ème chaîne (AZ) - est gère: Nationale
broadcastMode | 
collectionTitle | 20 heures le journal
corpus | 
descriptors | DET: alpinisme ; DET: accident ; DET: expédition ; DET: cascade ; DET: recherche ; DET: retour (rapatriement) ; DET: Népal ; DET: Himalaya :
duration | 00:05:27
eventDate | 
foreignOriginalVersion | 
generic | JOU Jacquier, Gilles ; PAR Koenig, Serge ; PAR Baud, Alain ;
genre | Journal télévisé ; Reportage :
id | 3129199001020
materialID | FPVD06071308 01
Credits

• Datasets:

• People:

• More info: http://newsml.cwi.nl
Literature


• Raphaël Troncy, Lynda Hardman, Jacco van Ossenbruggen and Michael Hausenblas: Identifying Spatial and Temporal Media Fragments on the Web. In W3C Video on the Web Workshop, San Jose (California) and Brussels (Belgium), December 2007.

• W3C Video on the Web Activity, April 2008 http://www.w3.org/2008/01/video-activity.
What are the messages?

- Features can be extracted and used to describe multimedia content (metadata)
- Multimedia presentations embody messages
- Media, structure and aesthetics all contribute to conveying the message
- The message can be made explicit (more metadata)
- Media content and metadata can be passed around and among systems
- We need to agree on how to do this (e.g. canonical processes; COMM; W3C working groups; IPTC)
- Users can be given much richer and more flexible access to (semantically annotated) content, but...
- we are still figuring out how to do this.
Thanks for your attention

http://www.cwi.nl/~media/samt08/