Semantic Multimedia Web

Material adopted from Raphaël Troncy & Lynda Hardman, CWI
Beyond Multimedia Databases

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

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

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*

Semantic Search and Presentation of Multimedia Content
- Link your data!
- *Searching and Browsing Multimedia Semantic Datasets with Cliopatria*
Overview of Canonical Processes

- Premeditate
- Create
- Annotate
- Package
- Construct Message
- Organize
- Query
- 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

Package

Create
CeWe Color PhotoBook Processes
CeWe Color PhotoBook Processes

Query:
- 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:
- Use largest images possible per page (show less 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

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
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
57 Claims, 16 Data, 4 Concessions, 3 Warrants, 1 Condition
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

```
diplomacy best solution
  \___\       \____ \ 
  support    contradict  support

war best solution
  \____\  
  war not solution
```
Result of Vox Populi Query

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

Publish

Distribute
Vox Populi Processes

ISWeb - Information Systems & Semantic Web
Steffen Staab
staab@uni-koblenz.de
Canonical Processes 101

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

Formalization of each process in UML diagrams

- Process
- Process artifacts
- Process actors
- External world artifacts
Process where initial ideas about media production are established

- Design a photo book of my last holidays for my family
- Create argument-based sequences of videos of interviews after September 11
Create Media Asset

Process where media assets are captured, generated or transformed
Process where annotation is created

Artifact Anchor

Any Process Artifact

Annotate

Artifact Annotation

Annotation Actor

Human Annotator

Computing Annotator
The annotation uses some controlled vocabularies:

- *Subject matter annotations of your photos*
- *Rhetorical annotations in Vox Populi*
Process where process artifacts are logically and physically packed

- **Any Process Artifact**
  - **Input given by a user**
    - **External world artifact**
- **Package**
  - **Input**
  - **Output**
- **Multimedia Package**
  - **Composite artifact**
- **Physical Package**
- **Logical Package**

contains
Process where a user retrieves a set of process artifacts based on a given query
Process where an 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*
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

![Diagram of process artifact relationships](image)
Publish

Process where final content and user interface is created

- **Document Structure** (from Organize)
- **Media Asset for Publication**
- **Annotation for Publication**
- **Publisher**
- **Published Document**
Distribute

Process where final interaction between end-users and produced media occurs
Canonical Processes Possible Flow

- Premeditate
- Create
- Annotate
- Package
- Organize
- Publish
- Construct Message
- Query
- Distribute
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

Special Issue on Canonical Processes of Media Production

http://www.ifi.uio.no/MMSJ/upcomming.html
http://www.cwi.nl/~media/projects/canonical/


Agenda

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

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*

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

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

ISO

W3C

ISO

W3C
ISO standard since December of 2001
Main components:
- Descriptors (Ds) and Description Schemes (DSs)
- DDL (XML Schema + extensions)
Concern all types of media

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
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></td>
<td>ABC</td>
<td>None</td>
<td>None</td>
<td>DOLCE</td>
</tr>
<tr>
<td>Complexity</td>
<td>OWL-Full</td>
<td>OWL-DL</td>
<td>OWL-DL</td>
<td>OWL-DL</td>
</tr>
<tr>
<td>Coverage</td>
<td>MDS+Visual</td>
<td>MDS+CS</td>
<td>All</td>
<td>MDS+Visual</td>
</tr>
<tr>
<td>Applications</td>
<td>Digital Libraries</td>
<td>Digital Libraries</td>
<td>Digital Rights</td>
<td>MM Analysis</td>
</tr>
</tbody>
</table>
Common Scenario

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

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

The "Big Three" at the Yalta Conference (Wikipedia)
Common Scenario: SW 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
- Link to knowledge on the Web

```prefix
:Reg1 foaf:depicts dbpedia:WinstonChurchill
dbpedia:Churchill rdfs:label "Winston Churchill"
dbpedia:Churchill rdf:type foaf:Person
```
Hunter's MPEG-7 Ontology

The Big Three at the Yalta Conference

mpeg7:image

mpeg7:depicts

mpeg7:spatial_decomposition

mpeg7:StillRegion

mpeg7:depicts

mpeg7:Polygon

mpeg7:Coords

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

mpeg7:StillRegion

mpeg7:depicts

mpeg7:DominantColor

rgb(25,255,255)

mpeg7:StillRegion

mpeg7:depicts

dbpedia:Churchill

The Big Three at the Yalta Conference


mpeg7:image
mpeg7:Title
contentString

mpeg7:SpatialDecomposition
mpeg7:SubRegion
mpeg7:SpatialMask
mpeg7:Polygon
mpeg7:Coord
5 25 10 20 15 10 10 5 15"^^xsd:string

mpeg7:MediaLocator

mpeg7:CreationInformation
mpeg7:Creation

mpeg7:RelatedMaterial
dbpedia:Churchill

mpeg7:StillRegion

mpeg7:MediaURI

DS-MIRF MPEG-7 Ontology
Rhizomik MPEG-7 Ontology


mpeg7:MediaLocator

mpeg7:SegmentType

rdf:type

dbpedia:Churchill

mpeg7:Semantic

mpeg7:SubRegion

mpeg7: SpatialMask

mpeg7: Spatial decomposition

mpeg7: CreationInformation

mpeg7: Title

The Big Three at the Yalta Conference

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

Link with domain semantics
- Hunter: ABC model + mpeg7:depicts relationship
- DS-MIRF: Domain ontologies needs to subclass the general MPEG-7 categories
- Rhizomik: Use the 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:WinstonChurchill
dbpedia:Churchill rdfs:label "Winston Churchill"
dbpedia:Churchill rdf:type foaf:Person
```
Scenario: Video

A history of G8 violence (video) (© Reuters)

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

:Seq1 foaf:depicts dbpedia:34th_G8_Summit
:Seq4 foaf:depicts dbpedia:EU_Summit
geo:Heiligendamn skos:broader geo:Germany
Research Problem

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

Multimedia objects are complex
- Compound information objects, fragment identification

Semantic annotation
- Subjective interpretation, context dependent

Linked data principle
- Open to reuse existing knowledge

⇒ MPEG-7
⇒ D&S | OIO
⇒ RDF
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
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
    - digital-data
  - multimedia-dat

- information-object
  - processing-role
    - output-role
    - input-role
  - segmentation-algorithm

- role
  - method
    - algorithm
      - segment-decomposition
    - setting

MPEG-7
COMM: Annotation Pattern

MPEG-7
Domain Ontologies
COMM: Modules

- Annotation Pattern
- Decomposition Pattern

Multimedia Knowledge (COMM)
Example 1: Fragment Identification


dns:realized-by

core:image-data

dns:plays

dns:played-by

dns:defines

dns:defines

loc:region-locator-descriptor

loc:spatial-mask-role

loc:bounding-box

5 25 10 20 15 15 10 10 5 15

xsd:string

data:has-rectangle
Example 1: Region Annotation


dns:realized-by

core:image-data
dns:setting

dns:plays

dns:played-by

dns:defines

dns:defined-by

data:has-rectangle

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

foaf:Person

dns:setting

core:semantic-annotation

dns:defines

core:semantic-label-role

dns:played-by

dns:played-by


dns:defines

core:semantic-label-role

dns:played-by

dns:played-by


dns:realized-by

core:image-data
dns:setting

dns:plays

dns:played-by

dns:defines

dns:defined-by

data:has-rectangle

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

foaf:Person

dns:setting

core:semantic-annotation

dns:defines

core:semantic-label-role

dns:played-by

dns:played-by

Example 2: Fragment Identification


dns:realized-by

core:image-data

dns:plays

dns:played-by

dns:defines

dns:defines

data:has-time

loc:media-time-descriptor

loc:temporal-mask-role

"1:21"^^xsd:time

ISWeb - Information Systems & Semantic Web
Steffen Staab
staab@uni-koblenz.de
Example 2: Sequence Annotation

dns:realized-by
dns:setting
dns:plays
dns:played-by
dns:defines
data:has-time
"1:21"^^xsd:time
data:has-time
loc:media-time-point
dns:defines
loc:media-time-descriptor
dns:played-by
loc:temporal-mask-role
dns:plays
core:semantic-annotation
dns:defines
tgn:Sweden
skos:broader
tgn:Gothenburg
tgn:Sweden
skos:broader
tgn:Gothenburg

core:image-data
dns:defines
loc:temporal-mask-role
core:semantic-label-role
dns:played-by
core:semantic-label-role
dns:played-by
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 meta-data at runtime
KAT Annotation Tool
Evaluation

Applied Domains
- Knowledge management for multimedia documents
- Driving multimedia analysis process
- Generate new interfaces for browsing multimedia content

Scalability
- 4 minutes video, TRECVid metadata expressed in COMM
  - 250 K statements
- Reasoning in large scale applications


Agenda

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

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*

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

Annotate the content (interpretation)
Elephant, Ganesh, Thailande, Holidays, Chiang Mai

Link to knowledge on the Web

```
<rdf:Description rdf:about="Ganesh.jpg">
  <dc:title>An image of the Elephant Ganesh</dc:title>
  <dc:creator>Raphaël Troncy</dc:creator>
</rdf:Description>
```

:img foaf:depicts dbpedia:Ganesh
dbpedia:Ganesh rdfs:label "Vinayaka"
dbpedia:Ganesh skos:altlabel "Ganapati"
dbpedia:Ganesh rdf:type wn:synset-Deities-noun-1
dbpedia:Ganesh owl:sameas wn:synset-Ganesh-noun-1
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)

http://richard.cyganiak.de/2007/10/lod/
DBpedia is a community effort to:

- extract structured "infobox" information from Wikipedia
- interlink DBpedia with other datasets on the Web
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
Automatic Links Among Open Datasets

Processors can switch automatically from one to the other ...
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
This cultural search engine will give you access to artworks from several museum collections. Type a keyword, for example, Derrain, calligraphy, or 1967.
Semantic Browsing of Multimedia News

Goal:
- Provide an environment for searching and browsing contextualized multimedia news information

Method:
- Semantic processing of multimedia news items
- Link news items with knowledge on the web

Datasets:
- News stories: Jun/Jul 2006 (en/fr) newsfeed, AFP ± 90,000 items
- Photos: 2006 football world cup, AFP ± 2,500 items
- Video: Jun/Jul 2006 TV News (fr), INA ± 30 items
Problems

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

NAR Schema

NewsCodes

Controlled Vocabularies

Broadcaster Schema

FoAF + SKOS

alignment

ISWeb - Information Systems & Semantic Web
Steffen Staab
staab@uni-koblenz.de
Semantic Processing

Named Entity Recognition

NAR Ontology
NewsCodes
Thesaurus

Domain Ontologies

spout
Semantic Processing

Knowledge Assisted Analysis

Concept Detectors

NAR Ontology
NewsCodes
Thesaurus

Domain Ontologies
Semantic Processing

- Named Entity Recognition
- Linked Data
- Domain Ontologies
- NAR Ontology
- NewsCodes
- Thesaurus

ISWeb - Information Systems & Semantic Web
Steffen Staab
staab@uni-koblenz.de
This news search engine will give you access to news items kindly provided by AFP.
Type a keyword, for example: Amsterdam, Lyon or Zidane.
Future Work

Integrate the video browser in the interface
  - Metadata conversion and interoperability
  - Address temporal fragments of the video
  - Visualize videoclips in the interface

Enrich metadata with visual analysis
  - Apply K-Space concept detectors on visual media
  - Provide new dimensions (facets) for browsing the data
    - Ex: distinguish field images vs stadium and street images with a grass detector for the World Cup dataset

Evaluation, Evaluation, Evaluation …

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.