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
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
- Package
- Create
- Annotate
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

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

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

Vox Populi Processes

I am not a fan of military actions
I cannot think of a more effective solution

Publish

War has never solved anything
Two billions dollar bombs on tents

Distribute
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

Premeditate
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

Annotate
Process where annotation is created
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
  - Annotate
  - Artifact Annotation
  - Vocabulary
    - Semantic Annotate
    - Vocabulary Term
  - Semantic Artifact Annotation

Subject matter annotations of your photos:
- Our holiday was sporty, great weather and fun
- Create clash about whether war is a good thing
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

Publish

Process where final content and user interface is created

Distribute

Process where final interaction between end-users and produced media occurs

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.ifi.uio.no/MMSJ/upcoming.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

Multimedia Web Tagging Semantic Web

W3C Multimedia Semantics XG

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

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

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>None</td>
<td>None</td>
<td>DOLCE</td>
<td></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)

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
Draw a bounding box, a circle around a shape

Interpret the content

Link to knowledge on the Web

 localize a region

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

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

⇒ MPEG-7
⇒ D&S | OIO
⇒ RDF

Scenario: Image

Scenario: Video

A history of G8 violence (© Reuters)

Seq1 foaf:depicts dbpedia:34th_G8_Summit
dbpedia:34th_G8_Summit geo:Heiligendamm skos:broader geo:Germany

Seq4 foaf:depicts dbpedia:EU_Summit
dbpedia:EU_Summit

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

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

COMM: D&S / OIO Patterns

- 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

<table>
<thead>
<tr>
<th>D&amp;S / OIO</th>
<th>defines</th>
<th>satisfies</th>
</tr>
</thead>
<tbody>
<tr>
<td>description</td>
<td>information-object</td>
<td>role</td>
</tr>
<tr>
<td>processing-role</td>
<td>method</td>
<td>situation</td>
</tr>
<tr>
<td>structured-data-description</td>
<td>digital-data</td>
<td>algorithm</td>
</tr>
<tr>
<td>input-role</td>
<td>output-role</td>
<td>input-role</td>
</tr>
</tbody>
</table>

MPEG-7

COMM: Annotation Pattern

<table>
<thead>
<tr>
<th>D&amp;S / OIO</th>
<th>defines</th>
<th>satisfies</th>
</tr>
</thead>
<tbody>
<tr>
<td>description</td>
<td>information-object</td>
<td>role</td>
</tr>
<tr>
<td>processing-role</td>
<td>method</td>
<td>situation</td>
</tr>
<tr>
<td>structured-data-description</td>
<td>digital-data</td>
<td>algorithm</td>
</tr>
<tr>
<td>input-role</td>
<td>output-role</td>
<td>input-role</td>
</tr>
</tbody>
</table>

MPEG-7
COMM: Semantic Pattern

Example 1: Fragment Identification

Example 1: Region Annotation
Example 2: Fragment Identification

Example 2: Sequence Annotation

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, TREC Vid metadata expressed in COMM
  - 250 K statements
- Reasoning in large scale applications

**Literature**


**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, Thailand, Holidays, Chiang Mai
Link to knowledge on the Web
- An image of the Elephant Ganesh
- Raphaël Troncy

```xml
<rdf:Description rdf:about="Ganesh.jpg">
  <dc:title>An image of the Elephant Ganesh</dc:title>
  <dc:creator>Raphaël Troncy</dc:creator>
</rdf:Description>
```
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
DBpedia is a community effort to:
• extract structured "infobox" information from Wikipedia
• interlink DBpedia with other datasets on the Web

DBpedia
Extracting Infobox Data

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

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

Semantic Processing

- Named Entity Recognition
- Domain Ontologies

- Knowledge Assisted Analysis
- Concept Detectors
- Domain Ontologies

- Linked Data
- RDF
- Linked Open Data
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 ...

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.