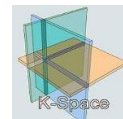


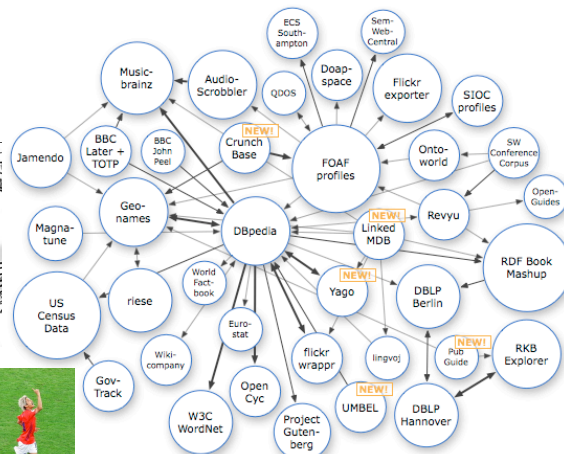
Multimedia lifecycle

Lynda Hardman

CWI, Interactive Information Access
UvA, Multimedia Interaction



Motivation



Creating a semantic media web

- Allocate URI to media asset
- Attach metadata using RDF

But...

- want to attach metadata to parts of a media asset
- want to combine multiple (parts) of assets into new assets/presentations
- each media type/format needs a different player (whereas everyone can display text)

Summer School on Multimedia Semantics, Koblenz, August 2009

3

We don't even care about media!

- We want to enable
 - the processing of information-bearing content
 - of one or more media types
 - that can be interpreted by end users
- End-users are primarily interested in
 - the meaning conveyed by a combination of media assets
 - interacting further with the media
 - as part of complex search task
 - passing it on to someone else in media "chain"

Summer School on Multimedia Semantics, Koblenz, August 2009

4

We can use the (semantic) web

- *Web* enables the identification and delivery of units of information of different data types
- *Semantic web* enables the association of metadata with each identified unit/fragment
- To find and use them, we need mechanisms:
 - for identifying (part of) an individual media asset
 - for associating metadata with an identified fragment
 - that enable larger meaningful structures to be composed, identified and annotated

Summer School on Multimedia Semantics, Koblenz, August 2009

5

Really need...

- to be aware of the human aspects of multimedia
- multimedia assets are not created "*in vacuo*", but by someone for a specific purpose
- more than the information "expressed" by the media asset itself, e.g.
 - the creator and the intended purpose
 - provenance also important on the semantic web, where a multimedia presentation may be composed of assets published by many different sources
 - the context in which the media asset was created (e.g. before or after Obama was elected)

Summer School on Multimedia Semantics, Koblenz, August 2009

6

Why are you here?

- To do just this...
- To understand the context in which information is used
- Learn about different techniques that already exist
- Understand how far media is part of the semantic web
- And most importantly, what still needs to be done to get it to work!

Summer School on Multimedia Semantics, Koblenz, August 2009

7

This talk

Understand Multimedia Applications Workflow

CeWe Color Photo Book creation application

Vox Populi argumentation-based video sequences generation

Canonical Processes of Media Production

8

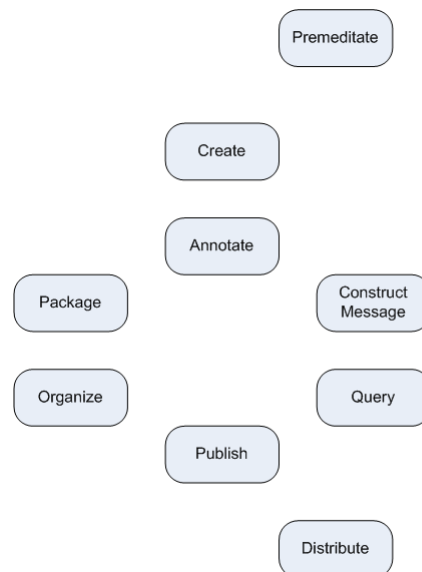
Understanding Multimedia Applications Workflow

- Identify and define a number of canonical processes of media production
- Community effort
 - 2005: [Dagstuhl seminar](#)
 - 2005: ACM MM Workshop on [Multimedia for Human Communication](#)
 - 2008: Multimedia Systems Journal Special Issue (core model and companion system papers)
editors: Frank Nack, Zeljko Obrenovic and Lynda Hardman



9

Overview of Canonical Processes



10

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

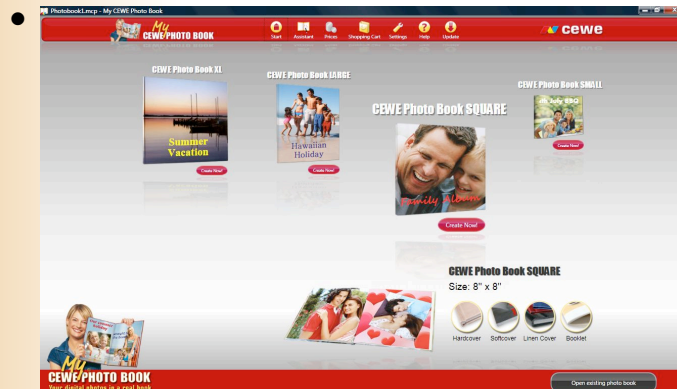
<http://www.cewe-photobook.com>

11

CeWe Color PhotoBook Processes

- My winter ski holidays with my friends

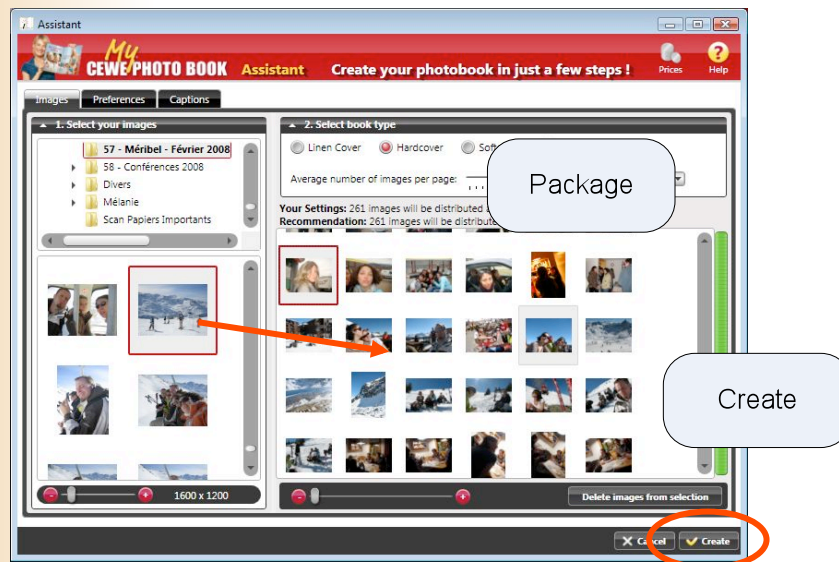
Premeditate



Construct
Message

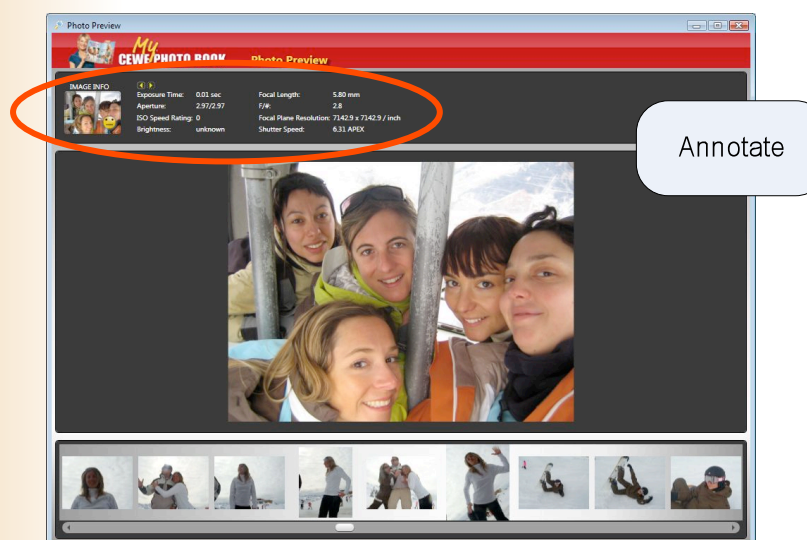
12

CeWe Color PhotoBook Processes



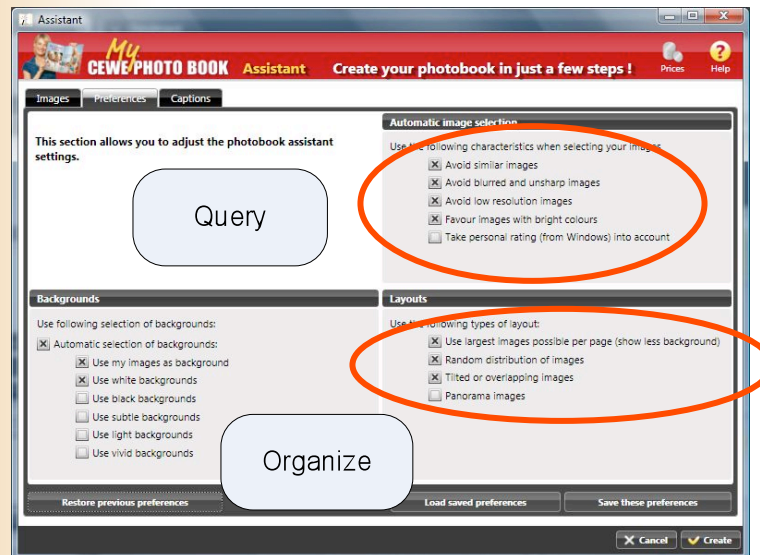
13

CeWe Color PhotoBook Processes



14

CeWe Color PhotoBook Processes



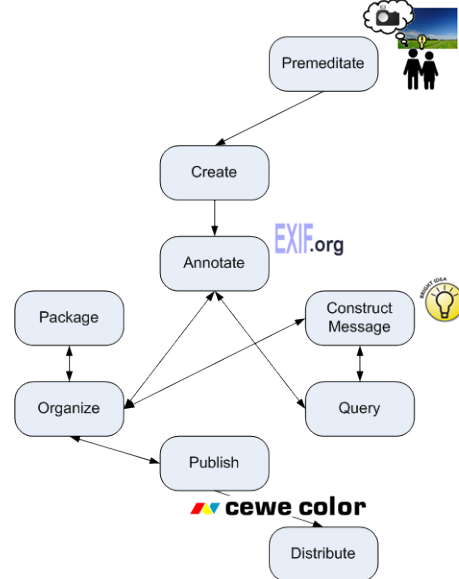
15

CeWe Color PhotoBook Processes



16

CeWe Color PhotoBook Processes



17

Generating video documentaries from annotated media repositories

Stefano Bocconi, Frank Nack
CWI Amsterdam
The Netherlands
Contact: stefano.bocconi@gmail.com



Talk Outline

- Motivation
 - Example
 - Scenarios
 - Technical details
 - Annotations
 - Editing Process
 - Conclusions
-

19

Video Documentaries on the Web

- Traditional video authoring: there is only one final version, what is shown is the choice of the author/editor
 - Proposed video authoring:
 - Annotate the video material semantics
 - Show automatically what the user asks to see, using presentation forms a film editor would use
-

20

Video material

- Focus on video interviews about controversial issues
- **Interview with America** video footage with interviews and background material about the opinion of American people after 9-11
www.interviewwithamerica.com

21

Example: *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..."

22

What do you think of the war in Afghanistan?

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

23

The annotations

- ☐ Rhetorical
 - Rhetorical Statement (mostly verbal, but visual also possible)
 - Argumentation model: Toulmin model
- ☐ Descriptive
 - Question asked
 - Interviewee (social)
 - Filmic *next slide*

24

Filmic annotations

Continuity, e.g.

- lighting conditions
- background sound
- gaze direction of speaker
left, centre, right
- framing continuity
close-up, medium shot, long shot
- camera movement
none, pan left/right, shaking, tilt up/down,
zoom in/out

25

Statement encoding

- Statement formally annotated:
 - <subject> <modifier> <predicate>
 - E.g. “**war best solution**”
- A thesaurus containing:
 - Terms (155)
 - Relations between terms:
similar (72), *opposite* (108),
generalization (10), *specialization* (10)
 - E.g. **war** *opposite* **diplomacy**

26

Connect statements

- ❑ Using the thesaurus, generate related statements and query the repository
"war best solution",
"diplomacy best solution",
"war not solution"
- ❑ Create a **graph** of related statements
 - nodes are the statements
(corresponding to video segments)
 - edges are either *support* or *contradict*

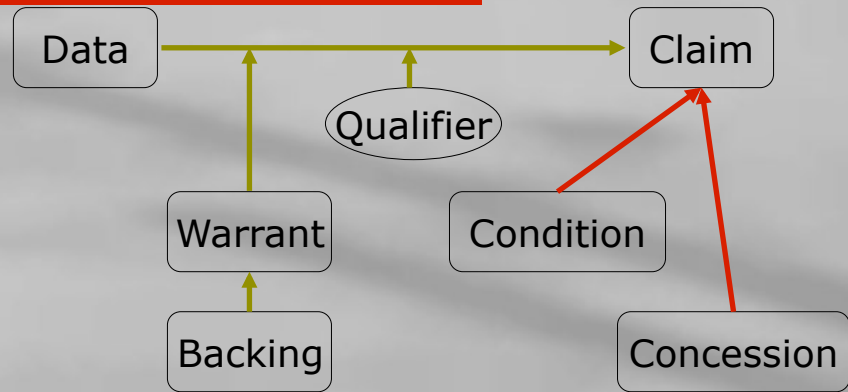
27

Semantic Graph



28

Toulmin model



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

29

Analysis of the Example

Two billions dollar bombs on tents



Claim

contradict



Claim

I cannot think of a more effective solution

weaken

Concession

I am not a fan of military actions



Claim

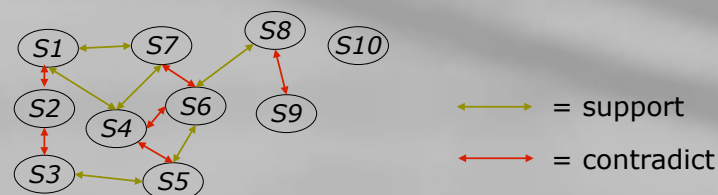
support

War has never solved anything

30

Facts and features

- Annotations: 1 hour annotated, 15 interviews, 60 interview segments, 120 statements
- Partially **tunable**: examining the Segment graph gives feedback on the quality of the annotations and the thesaurus



31

Controlling the Bias

- Video documentaries are not neutral account of reality: the selection and editing of the footage expresses a point of view
- Editing strategy:
 - Balanced
 - Pro opinion X
 - Against opinion X
- We use:
 - Logos (the statements)
 - Ethos (based on user profile)
 - Film editing (e.g. framing, gaze)

32

Vox Populi interface

The interface is divided into several sections for configuring a video interview:

- Question:** A list of questions to choose from, including "Why did they do what they did?", "What do you think of the casualties among civilians?", "What do you think of the Afghanistan war?", "What are the consequences of the war?", "What are the roots of the problem?", and "What do you think about the Anthrax?".
- Interviewee:** A dropdown menu showing "Cameron Parking Guard at Stamford" and "Lawyer in Harvard".
- Opinion:** A dropdown menu showing "War in Afghanistan - Pro".
- Position:** A label indicating the role of the interviewee.
- First Character:** A set of dropdown menus for character attributes: Age (Middleage, Old, Teenager, Young), Education (HighEducated, LowEducated, MediumEducated), Employment (HighIncomeJob, LowIncomeJob, MiddleIncomeJob, Retired, Student), GeoLocation (NotUSA, USA), Race (AmericanIndian, Asian, Black, Hispanic, White), Religion (Atheist, Christian, Muslim), and Sex (Female, Male).
- Second Character:** A second set of dropdown menus for character attributes, identical to the first character's.
- Strategy:** Radio buttons for "None", "Create Clash", "Create Support", and "Vox Populi".
- Bandwidth:** Radio buttons for "Low Bandwidth", "Medium Bandwidth", and "High Bandwidth".
- Intercut:** Radio buttons for "True" and "False".
- Caption:** Radio buttons for "On (can cause problems)" and "Off".
- Buttons:** "Done" and "Reset" buttons at the bottom left.

Conclusions

- ❑ Automatic generation of video interviews augmented with supporting and/or contradicting material
- ❑ The **user** can determine the subject and the bias of the presentation
- ❑ The **documentarist** can add material and let the system generate new documentaries

Pointers & Acknowledgments

- This presentation and a Demo available at:

<http://www.cwi.nl/~media/demo/IWA/>

- This research was funded by the Dutch national ToKeN I²RP and CHIME projects.

35

Literature

- *Stefano Bocconi*
Vox Populi: generating video documentaries from semantically annotated media repositories
PhD thesis, Technical University of Eindhoven, 2006
<http://www.cwi.nl/~media/Theses/Bocconi/VoxPopuli.pdf>

36

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..."

37

Vox Populi Premeditate Process

- Analogous to the pre-production process in the film industry

- *Static* versus *dynamic* video artifact

Premeditate

- 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

38

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

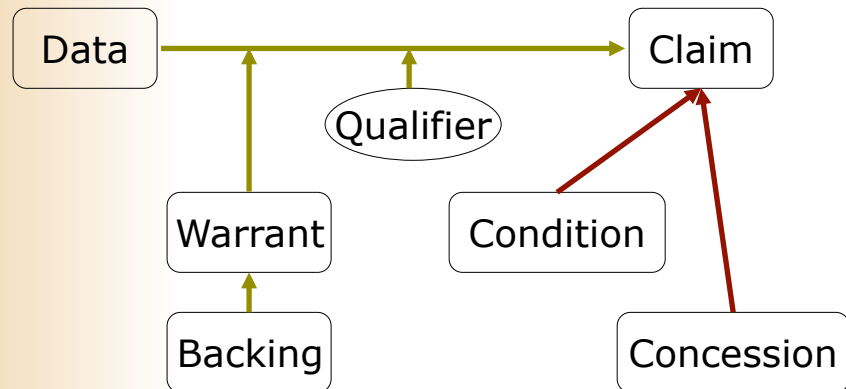
39

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**

40

Toulmin Model



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

41

Vox Populi Query Interface

Construct Message

Question: Why did they do what they did?
What do you think of the casualties on the civilians?
What do you think of the Afghanistan war?
What are the consequences of the war?
What are the roots of the problem?
What do you think about the Anthrax?

Interviewee: Cameroun Parking Guard at Stamford
Lawyer in Harvard

Opinion: War in Afghanistan - Pro

Position

Age	Education	Employment	GeoLocation	Race	Religion	Sex
Middleage	HighEducated	HighIncomeJob	NotUSA	AmericanIndian	Atheist	Female
Old	LowEducated	LowIncomeJob	USA	Asian	Christian	Male
Teenager	MediumEducated	MiddleIncomeJob		Black	Muslim	
Young		Retired		Hispanic		
		Student		White		

First Character

Age	Education	Employment	GeoLocation	Race	Religion	Sex
Middleage	HighEducated	HighIncomeJob	NotUSA	AmericanIndian	Atheist	Female
Old	LowEducated	LowIncomeJob	USA	Asian	Christian	Male
Teenager	MediumEducated	MiddleIncomeJob		Black	Muslim	
Young		Retired		Hispanic		
		Student		White		

Second Character

Strategy: ☒ None ☐ Create Clash ☐ Create Support ☐ Vox Populi

Bandwidth: ☐ Low Bandwidth ☐ Medium Bandwidth ☐ High Bandwidth

Intercut: ☐ True ☐ False

Caption: ☐ On (can cause problems) ☐ Off

Query

Done Reset

42

Vox Populi Organize Process

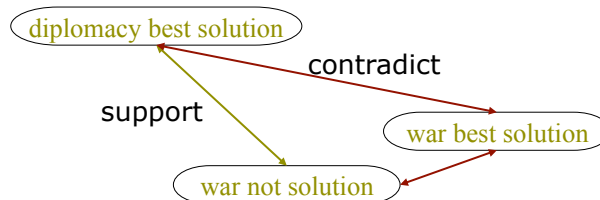
- Using the thesaurus, create a **graph** of related statements

- nodes are the statements
(corresponding to video segments)

"war best solution",
"diplomacy best solution",
"war not solution"

Organize

- edges are either *support* or *contradict*



43

Result of Vox Populi Query

**I am not a
fan of
military
actions**

**I cannot think
of a more
effective
solution**

Publish



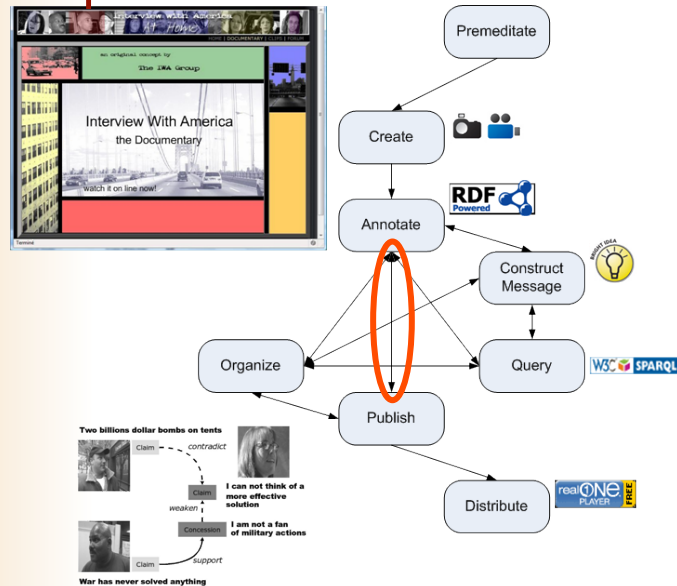
**War has
never
solved
anything**

**Two billions
dollar bombs
on tents**

Distribute

44

Vox Populi Processes



45

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)

46

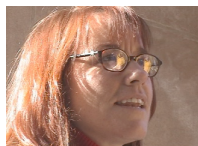
Premeditate

- 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

47

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



48

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...

49

Semantic Annotate

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



Subject

Modifier
thesaurus

Predicate
thesaurus

war

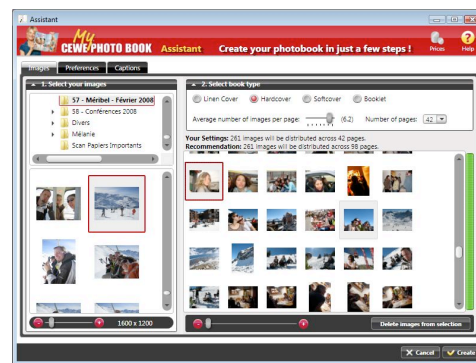
best

solution

50

Package

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



51

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)



52

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

53

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



54

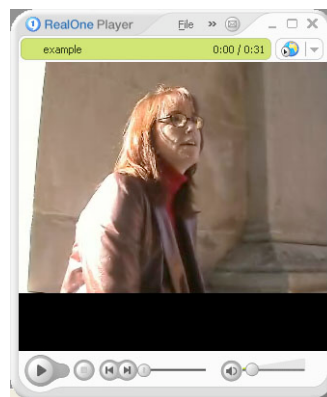
Publish

- 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

55

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



56

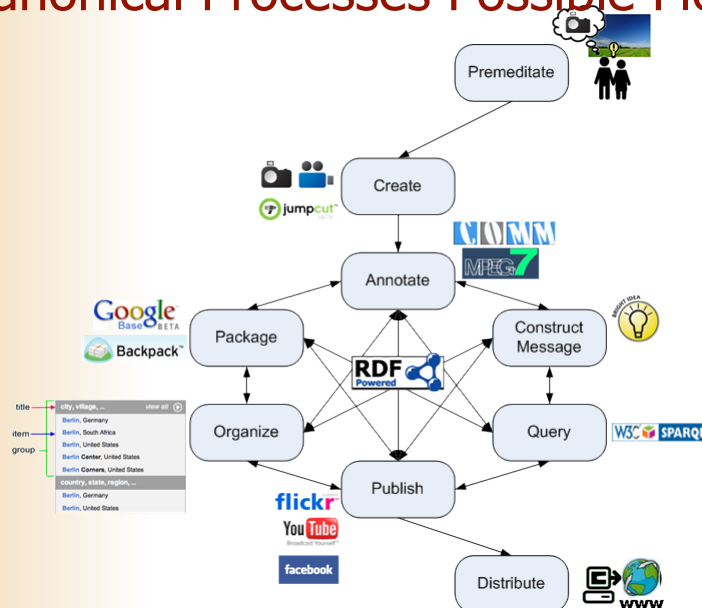
Canonical processes exercise

- Consider one of the following applications (or choose another – check with me first)
 - video surveillance
 - interactive media art
 - news
 - interactive digital drama
- Go through each of the processes and make brief notes for each in terms of your chosen application
 - premeditate, create, annotate, construct message, query, package, organise, publish, distribute

Summer School on Multimedia Semantics, Koblenz, August 2009

57

Canonical Processes Possible Flow



SAMT 2008 Tutorial: A Semantic Multimedia Web, 3 December 2008

58

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

59

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

Upper Ontology

**Model of Canonical Processes
of Media Production**

Models of Specific Media
Production Processes

Concrete Systems

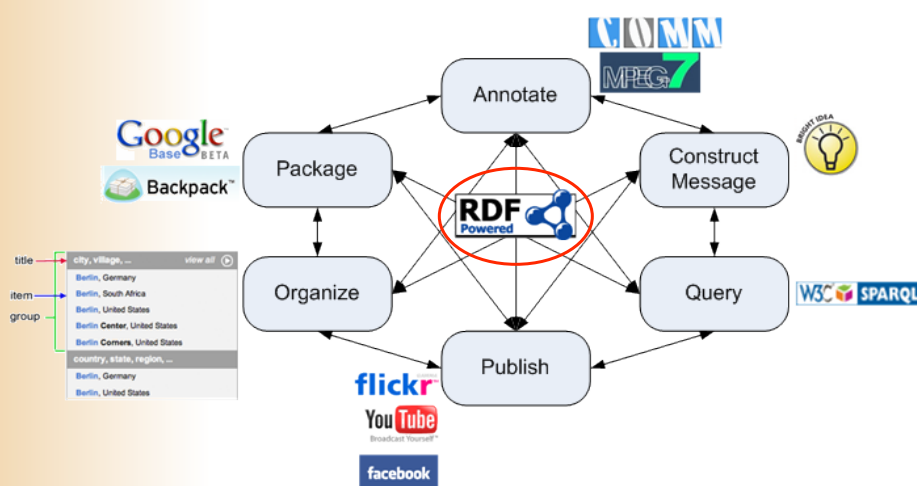
60

Literature

- Special Issue on Canonical Processes of Media Production
<http://www.springerlink.com/content/j0l4g337581652t1/>
<http://www.cwi.nl/~media/projects/canonical/>
- Lynda Hardman, Zeljko Obrenovic, Frank Nack, Brigitte Kerhervé and Kurt Piersol: *Canonical Processes of Semantically Annotated Media Production*. In [Multimedia Systems Journal](#), 2008
- Philipp Sandhaus, Sabine Thieme and Susanne Boll: *Canonical Processes in Photo Book Production*. In [Multimedia Systems Journal](#), 2008
- Stefano Bocconi, Frank Nack and Lynda Hardman: *Automatic generation of matter-of-opinion video documentaries*. In [Journal of Web Semantics](#), 6(2), p139-150, 2008.
- Lynda Hardman: *Canonical Processes of Media Production*. In [Proceedings of the ACM Workshop on Multimedia for Human Communication - From Capture to Convey \(MHC 05\)](#), November 2005.

61

Annotations: the link with COMM



62

Annotating an Image

Reg1



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
```

63

Annotating a Video



- Localize a region
 - Annotate the content
 - Tag: G8 Summit, Heiligendamn, 2007
 - Link to knowledge on the Web
- EU Summit, Gothenburg, 2001

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

64

Core Ontology on Multimedia - Mozilla Firefox

http://comm.semanticweb.org/

Search News RDFa Highlight Raphaël Troncy Mélanie CWI K-Space NewsML FP7, Call 3 W3C Conferences Planet RDF rammx (RDFa-deploy... ShapeShift.TV






Google Rechercher RS Mes favoris PaperClick Traduire Envoyer à Paramètres

COMM core ontology for multimedia

Home Ontology Examples Java API Papers

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](#)
- [Raphaël Troncy](#)
- [Richard Arndt](#)

Terminé

65

Acknowledgements



Raphaël Troncy, Eurecom

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