Semantic Annotation and Presentation of Multimedia Content

Raphaël Troncy
<Raphael.Troncy@cwi.nl>

CWI, Semantic Media Interfaces

Agenda

1. Image/Video: search and description problems
2. A Multimedia Semantic Web:
   - Introduction to RDF, RDFS and OWL
   - The Linked Data Principle
   - The return of MPEG-7
3. Generate Multimedia Presentations

Motivation

- Growing amount of digital multimedia content available on the Web
- Content difficult to search and reuse

Image search

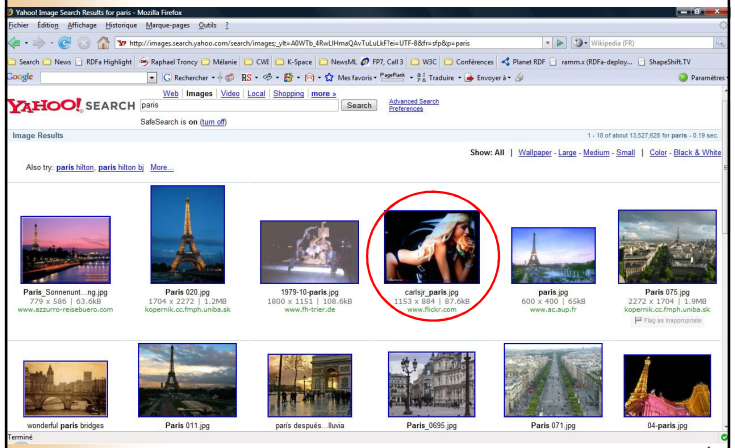
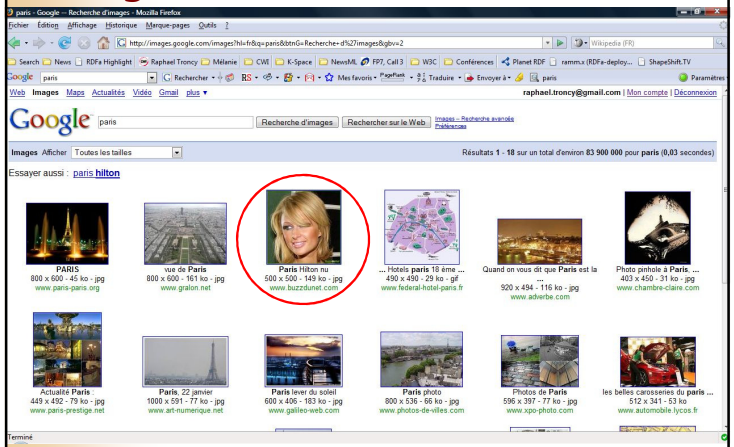
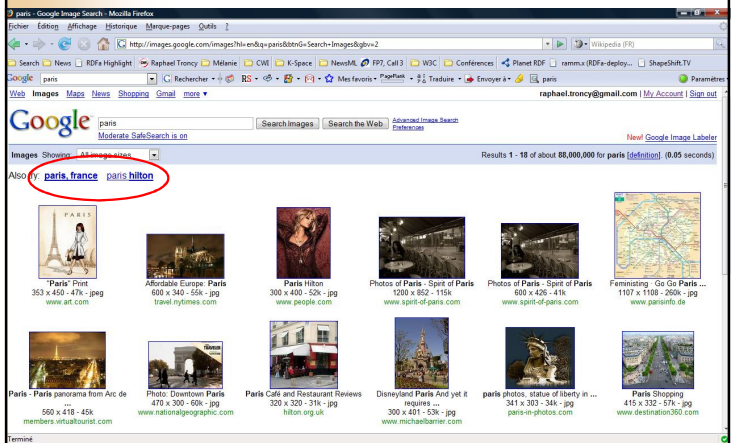


Image search



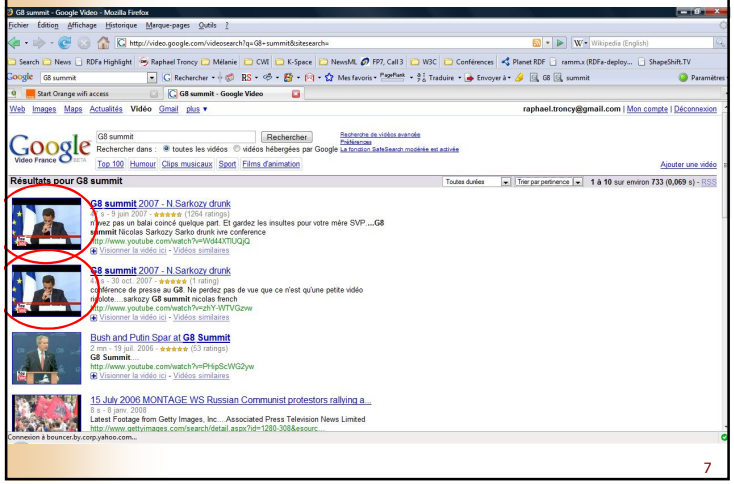
5

Image search



6

Video search



7

Image/Video indexing

- Techniques used by current search engines
  - search term occurs in the *filename* or in the *caption*
  - no semantics
- Image indexing: main problem
  - an image is not alphabetic: there is no countable discrete units, that, in combination will provide the meaning of the image
  - image descriptors are not given with the image: one needs to **extract** or **interpret** them
- Video indexing: additional problem
  - a video has additionally a temporal dimension to take into account
  - a video has *a priori* no discrete units neither (i.e. frames, shots, sequences cannot be absolutely defined)

8

Audio-Visual content

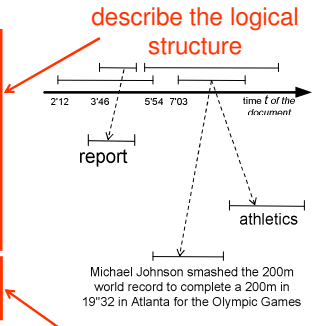
- The *audio-visual* document: some peculiarities
  - structured
  - spatio-temporal
  - composed of images

  } use of a textual description
- A three step process: (TV point of view)
  - ***identification*** of the content creator and the content provider: Dublin Core metadata, VRA core categories ...
  - ***structural decomposition*** in video segments corresponding to the logical structure of the program: time-code, spatial coordinates
  - ***semantic description*** of these segments: controlled vocabulary, thesaurus, free text annotation

9

Video description (TV point of view)

- Segmentation
  - locate and date some events
- Description
  - characterize each segment with an AV genre
  - characterize each segment with a general thematic
  - describe the scene (*who, when, where, what…*)



10

Full text search limitation

13 [**Indoor Set: 6th part**]
at 18:43:56:00 - 00:09:06:00. – Eurosport

In studio, the second part of the interview, from Nice, of Sandy CASAR by Jean René GODART about the Paris-Nice cycling race and a few sports news with pictures commented by Alexandre BOYON and Laurent PUYAT.

Q : Find all AV sequences of type dialog sequence with a rider and concerning any *cycling race with several stages*

- noisy answer: there are other *sports news* in the sequence
- incomplete answer: the interview was broadcasted in two parts and began in a previous sequence
- the query cannot be extended !

11

Image/Video summary

- Visual media are not self descriptive!
- Goal: find out the descriptors that describe their content
  - use a textual representation for the description / index
  - depends on the usage context of the media
- How to obtain these descriptors?
  - automatic analysis
    - segmentation; face/object detection; speech transcription; etc.
  - manual annotation
    - tagging; documentation
- How to represent these descriptors?
  - machine processable semantics

12

Video segmentation

- Allow a non linear access to the video
  - shot detection
  - scene segmentation (e.g. DVD chapters)
- Sudden transitive detection (« cut »)
  - based on similarity between consecutives images using color, motion, etc.
- Progressive transition detection
  - fade, wipe, etc.
  - based on statistical models of occurrences of various types of transitions

13

Shot segmentation example

- Good performance for « cut » detections:
  - 95% recall; 15% false detection
- Worse results for progressive detection such as fade

14

Object detection and recognition

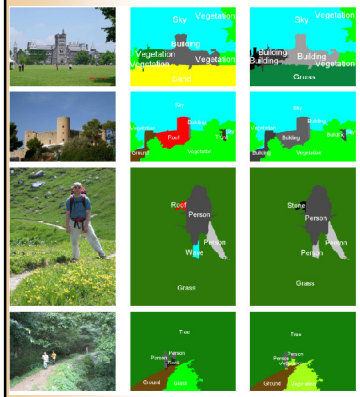
- Face/Object Detection
  - algorithms based on colour and face/object shape model
- Face/Object Recognition
  - put a name on a face/object
  - knowledge base of labelled faces/objects
  - probabilistic mapping rules
- Performance: efficient for front view faces
  - 80% recall, 3% false detection

15

Face detection examples

16

Object detection examples



- Left column: initial image
- Middle Column: Segmented image with initial labelling as a result of SVM-based classification
- Right Column: Final labelling, after the application of a Genetic Algorithm

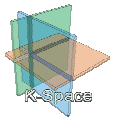


Image tagging

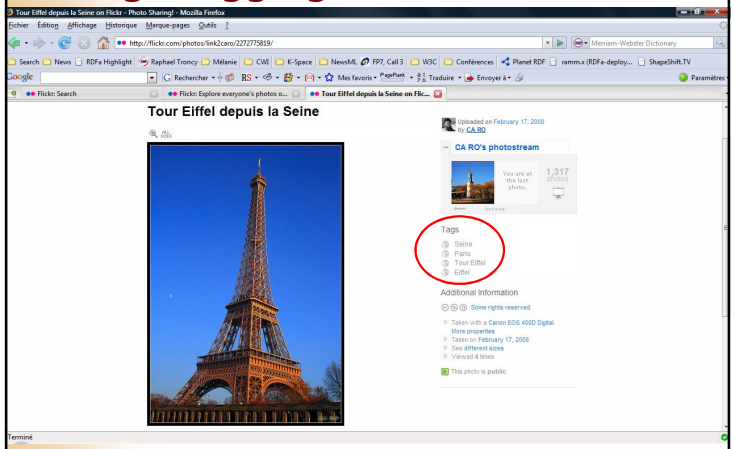


Image tagging: search

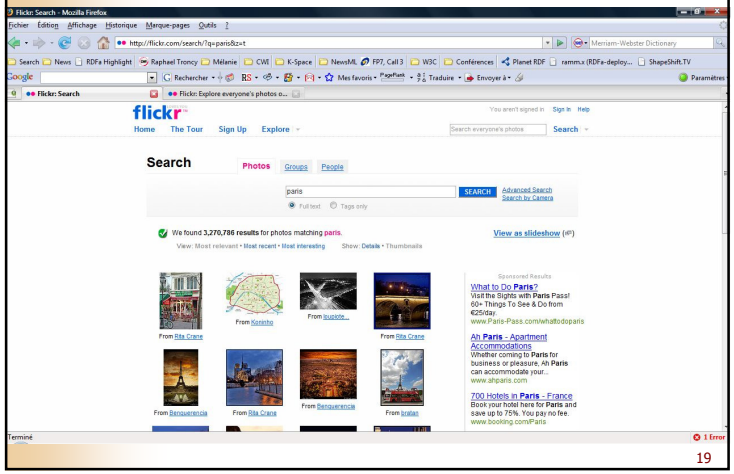
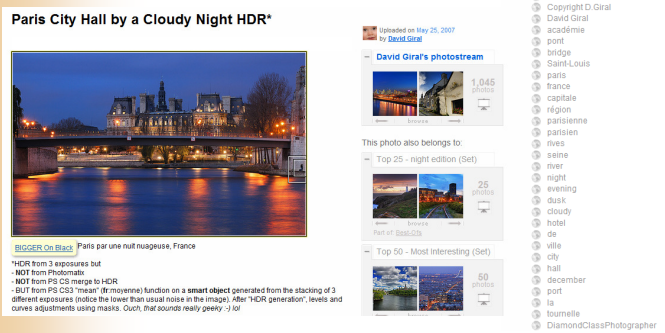
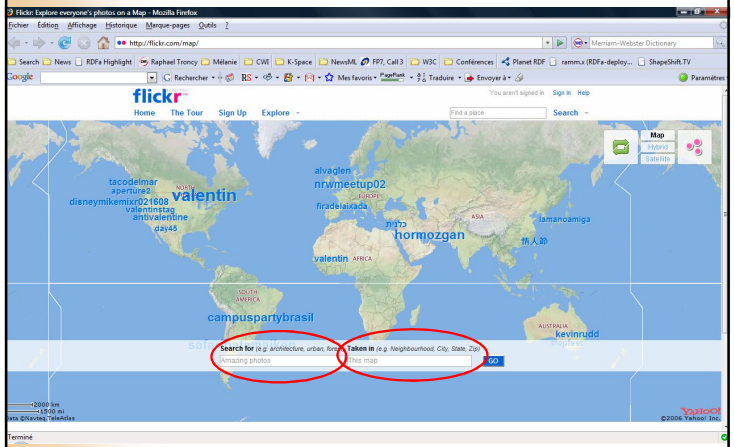


Image tagging: limitations

- tags are ambiguous: *Paris*
- tags are misspelled: *LA*, *Los Angeles*, *LosAngeles*
- tags are not structured

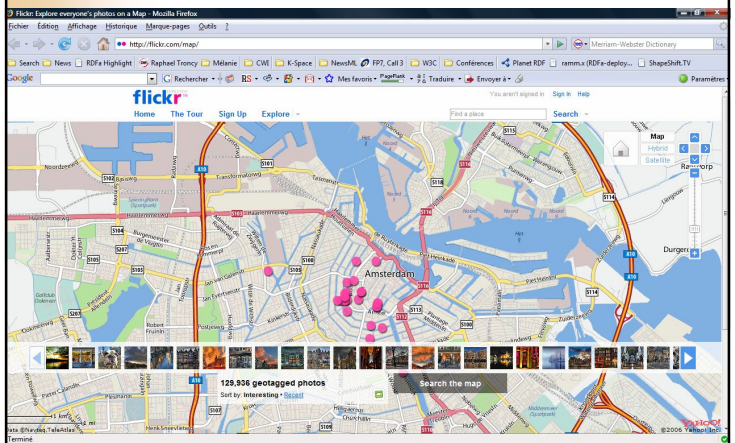


Geo-tagged images: machine tags



21

Geo-tagged images: machine tags



22

Multimedia description: summary

- Extract image descriptors (visual processing algorithms)
  - stay close to the physical nature of the media
  - machine learning: need learning data! (See: IBM Marvel)
  - limited performance in average situation
- Interpret the meaning of an image
  - usage context and culture dependent
  - manually assign tags: ambiguity, misspelling, error-prone
- Solution
  - structure / type the tags (see the machine tags)
  - give a ***formal*** meaning to the descriptions

23

Agenda

1. Image/Video: search and description problems
2. A Multimedia Semantic Web:
   - Introduction to RDF, RDFS and OWL
   - The Linked Data Principle
   - The return of MPEG-7
3. Generate Multimedia Presentations

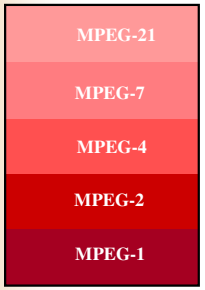
24

W3C Multimedia Semantics XG

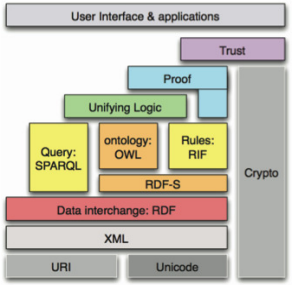
W3C Multimedia Semantics XG
http://www.w3.org/2005/Incubator/mmsem/

25

Multimedia: Description methods



ISO

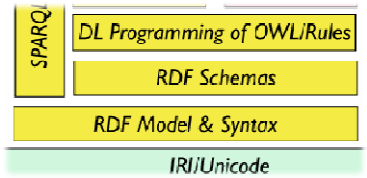


W3C

26

RDF

is the first layer of the
Semantic Web standards



27

RDF

stands for
Resource Description Framework

28

RDF

stands for

Resource: *pages, images, videos, ... everything that can have a URI*

Description: *attributes, features, and relations of the resources*

Framework: *model, languages and syntaxes for these descriptions*

29

RDF

is a triple model *i.e.* every piece of knowledge is broken down into

(subject , predicate , object)

30

take for instance the following piece of knowledge

31

image.jpg has for creator Raphael and depicts the elephant Ganesh

32

image.jpg has for creator Raphael
image.jpg depicts the elephant Ganesh

33

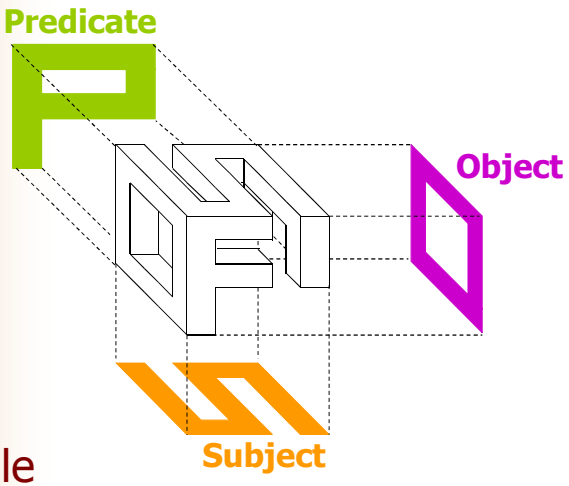
(image.jpg , creator , Raphael)
(image.jpg , depicts , Elephant Ganesh)

(subject , predicate , object)

34

in RDF the atoms of knowledge are triples of the form (subject,predicate,object)

35

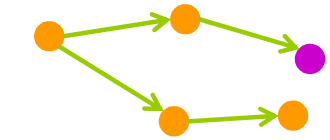


a triple
the RDF atom

36

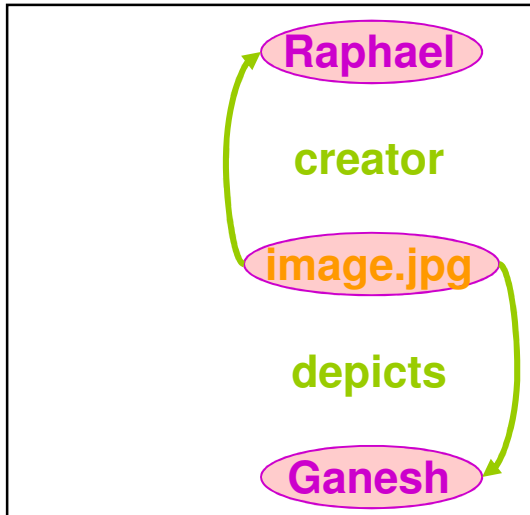
RDF is also a graph model
to link the descriptions of resources

37



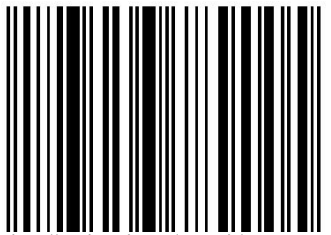
RDF triples can be seen as arcs
of a graph (vertex,edge,vertex)

38



39

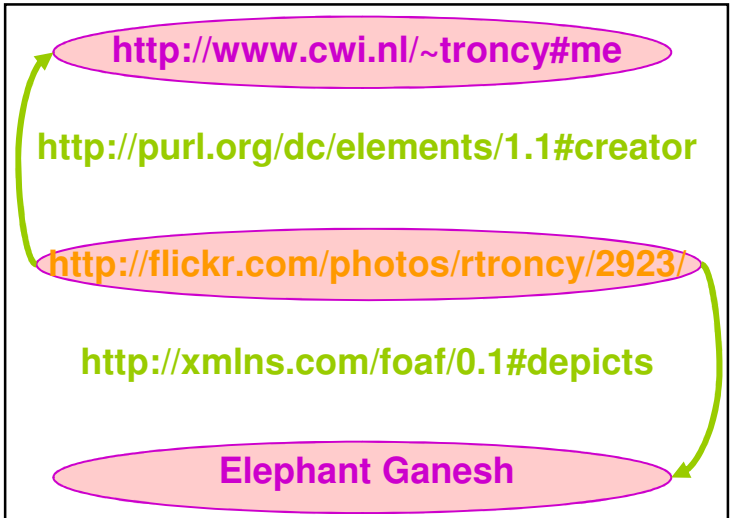
in RDF resources and properties are
identified by URIs



40

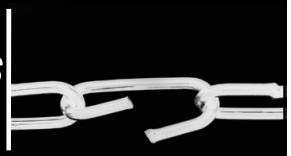
in RDF values of properties can also be literals i.e. strings of characters

41



42

RDF allows blank ***nodes***



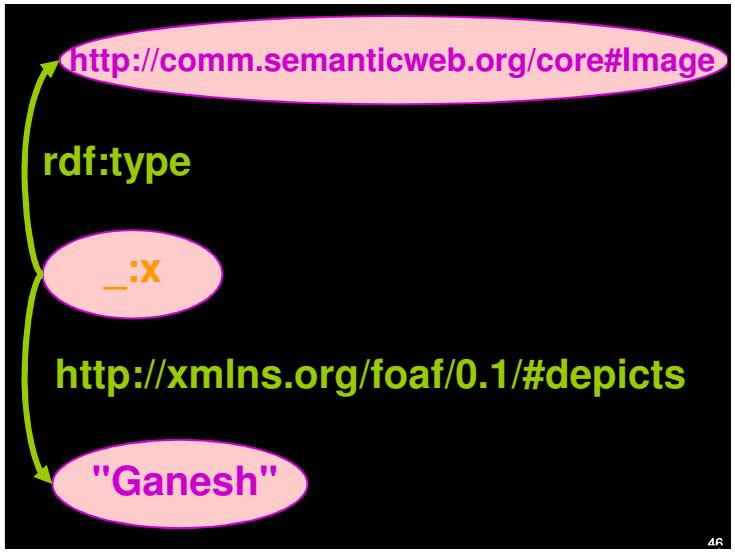
43

a resource may be
anonymous
i.e. not identified by a URI
and noted _:*xyz*

44

e.g.,
there exists an image about Ganesh

45



46

blank nodes
break
the graph, they cannot be reused

47

name
your resources and reuse existing names as much as possible

48

< RDF /> has an XML syntax

49

don't run
away

50

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-
  rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/" >

  <rdf:Description rdf:about=
  "http://flickr.com/photos/rtroncy/2923">
     <dc:creator rdf:resource=
         "http://www.cwi.nl/~troncy#me"/>
     <foaf:depicts>Ganesh</foaf:depicts>
  </rdf:Description>

</rdf:RDF>
```

51

it's only for
machines

52

open-world assumption

as opposed to the closed world assumption of classical systems

53

in short: the ***absence*** of a triple is *not* significant

54

(image.jpg , creator , Raphael)

doesn't mean image.jpg has one creator

55

(image.jpg , creator , Raphael)

means image.jpg has at least one creator

56

if you have no other triples giving authors it does not ***mean*** they are not true.

57

RDFS stands for RDF Schema

58

RDFS provides primitives to write lightweight schemas for RDF triples

59

a little drop of semantics goes a **long way**

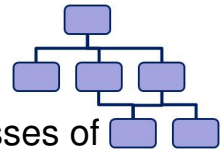
60

RDFS provides primitives to...

... define the vocabulary used in triples

... define elementary inferences

61



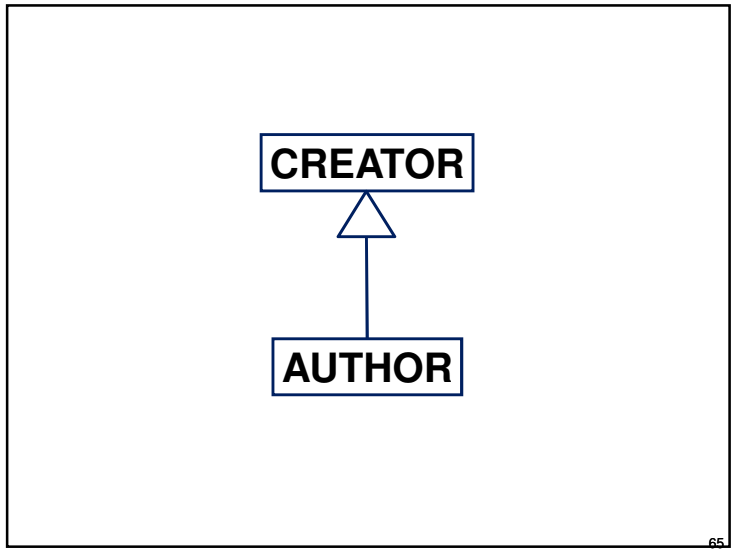
RDFS to define classes of resources and organize their hierarchy

62

63

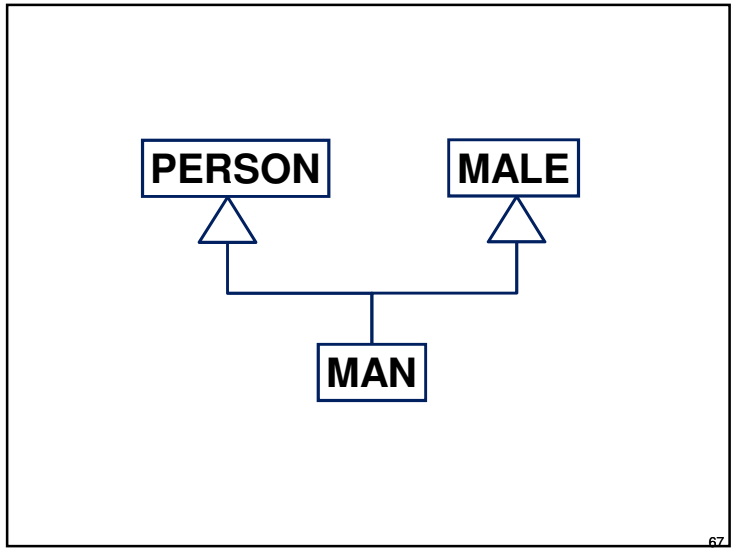
RDFS to define relations between resources and organize their hierarchy

64



RDFS allows for multiple inheritance for classes and properties

66



RDFS relations have a signature

RDFS relations have a signature

... the domain is the type of the resource the relation starts from.

... the range is the type of the resource the relation ends to.

69

RDFS relations with several...

... domains mean all domains apply.

... ranges mean all ranges apply.

Ranges and domains are optional

70

looks like object programming, **but**

71

properties

are first class citizens,

they are not defined inside classes,

they have their own hierarchy.

72

no overwriting

in particular a property can not be refined for sub classes of its range or domain.

73

multi-instantiation

a resource can have several types,
it can be the instance of several classes
like multiple lights, facets on a resource.

74

RDFS provides primitives to...

... give labels ...

... give comments ...

... for classes and properties

75

RDFS using the XML syntax for **RDF...**

76

```
<rdf:RDF xml:base="http://inria.fr/2005/humans.rdfs"
        xmlns:rdf ="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
        xmlns="http://www.w3.org/2000/01/rdf-schema#>



<Class rdf:ID="Man">
 <subClassOf rdf:resource="#Person"/>
 <subClassOf rdf:resource="#Male"/>
 <label xml:lang="en">man</label>
 <comment xml:lang="en">a male person</comment>
</Class>
```

77

```
<rdf:Property rdf:ID="hasMother">
 <subPropertyOf rdf:resource="#hasParent"/>
 <range rdf:resource="#Female"/>
 <domain rdf:resource="#Human"/>
 <label xml:lang="en">has for mother</label>
 <comment xml:lang="en">a female parent</comment>
</rdf:Property>



</rdf:RDF>
```

78

RDFS semantics : standard inference rules to derive additional triples from known statements.

79

IF (c_2, subClassOf, c_1)
AND (x, type, c_2)
THEN (x, type, c_1)

example of type propagation

IF (Man, subClassOf, Person)
AND (Tom, type, Man)
THEN (Tom, type, Person)

80

IF (p_2, subPropertyOf, p_1)
AND (x, p_2 , y)
THEN (x, p_1 , y)

example of property propagation

IF (author, subPropertyOf, creator)
AND (Tom, author, Report12)
THEN (Tom, creator, Report12)

81

IF (c_2, subClassOf, c_1)
AND (c_3, subClassOf, c_2)
THEN (c_3, subClassOf, c_1)

example of subClass transitivity

IF (Person, subClassOf, Animal)
AND (Man, subClassOf, Person)
THEN (Man, subClassOf, Animal)

82

IF (p_2, subPropertyOf, p_1)
AND (p_3, subPropertyOf, p_2)
THEN (p_3, subPropertyOf, p_1)

example of subProp transitivity

IF (parent, subPropertyOf, ancestor)
AND (father, subPropertyOf, parent)
THEN (father, subPropertyOf, ancestor)

83

IF (p_1, domain, c_1)
AND (x, p_1, y)
THEN (x, type, c_1)

example of domain inference

IF (author, domain, Human)
AND (Tom, author, Report12)
THEN (Tom, type, Human)

84

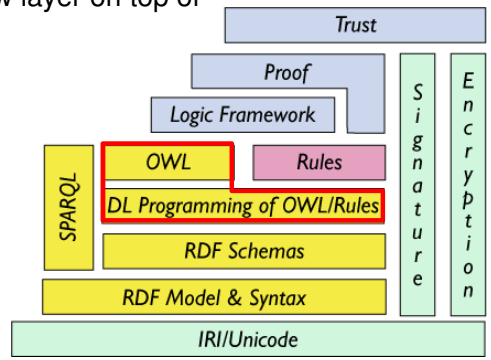
IF (p_1, range, c_1)
AND (x, p_1, y)
THEN (y, type, c_1)

example of range inference

IF (author, range, Document)
AND (Tom, author, Report12)
THEN (Report, type, Document)

85

- OWL is a new layer on top of RDFS
  - OWL Lite
  - OWL DL
  - OWL Full



- Actually, OWL is a restriction of RDF
  Classes ≠ Properties ≠ Individuals

86

RDF Schema Features:

- Class (Thing, Nothing)
- rdfs:subClassOf
- rdf:Property
- rdfs:subPropertyOf
- rdfs:domain
- rdfs:range
- Individual

(In)Equality:

- equivalentClass
- equivalentProperty
- sameAs
- differentFrom
- AllDifferent
- distinctMembers

Property Characteristics:

- ObjectProperty
- DatatypeProperty
- inverseOf
- TransitiveProperty
- SymmetricProperty
- FunctionalProperty
- InverseFunctionalProperty

Property Restrictions:

- Restriction
- onProperty
- allValuesFrom
- someValuesFrom

Restricted Cardinality:

- minCardinality (only 0 or 1)
- maxCardinality (only 0 or 1)
- cardinality (only 0 or 1)

Header Information:

- Ontology
- imports

Class Intersection:

- intersectionOf

Datatypes

- xsd datatypes

Versioning:

- versionInfo
- priorVersion
- backwardCompatibleWith
- incompatibleWith
- DeprecatedClass
- DeprecatedProperty

Annotation Properties:

- rdfs:label
- rdfs:comment
- rdfs:seeAlso
- rdfs:isDefinedBy
- AnnotationProperty
- OntologyProperty

87

Class Axioms:

- oneOf, dataRange
- disjointWith
- equivalentClass (applied to class expressions)
- rdfs:subClassOf (applied to class expressions)

Boolean Combinations of Class Expressions:

- unionOf
- complementOf
- intersectionOf

Arbitrary Cardinality:

- minCardinality
- maxCardinality
- cardinality

Filler Information:

- hasValue

88

Semantic Web Tutorials

RDF in a nutshell

- Tutorial (credits @ Fabien Gandon)

RDFS in a nutshell

- Tutorial (credits @ Fabien Gandon)

Useful Resources

- http://www.w3.org/2001/sw/

- http://www.w3.org/People/Ivan/CorePresentations/

A Giant Graph Open to the World

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

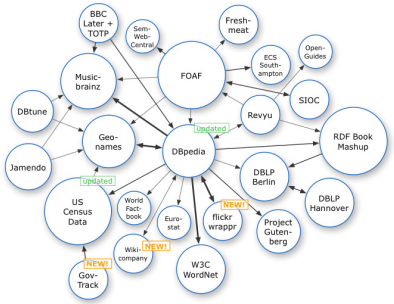
- Annotate the content (interpretation)
  Elephant, Ganesh, Thailande, Holidays, Chiang Mai
- Link to knowledge on the Web

```
: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

In a nutshell

- "Expose" open datasets in RDF
- Set RDF links among the data items for different datasets
- Over 2 billion triples, 3 millions links (November 2007)



DBpedia

DBpedia is a community effort to

- extract structured ("infobox") information from Wikipedia
- interlink the DBpedia dataset with other datasets on the Web

UNIVERSITÄT LEIPZIG Freie Universität Berlin OPENLINK SOFTWARE

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:population_metro "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)

93

Automatic Links Among Open Datasets

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

DBpedia

Geonames

```
<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"
 ...
```

Processors can switch automatically from one to the other …

94

How to Get RDF Data?

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

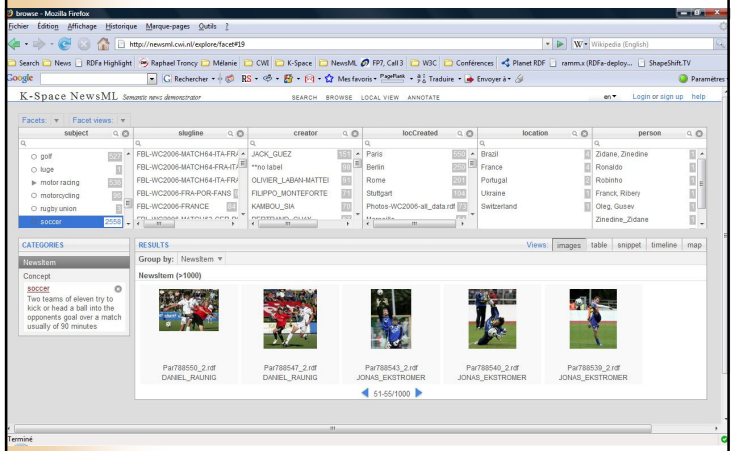
- service to get RDF from flickr images
  http://www.kanzaki.com/works/2005/imgdsc/flickr2rdf
- service to get RDF from XMP
  http://www.ivan-herman.net/cgi-bin/blosxom.cgi/WorkRelated/SemanticWeb/xmpextract.html

95

Take Home Message

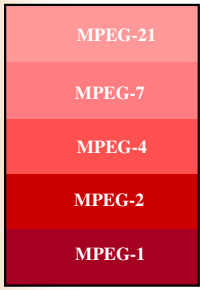
96

Demonstration

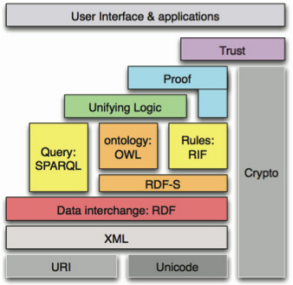


97

Multimedia: Description methods



ISO

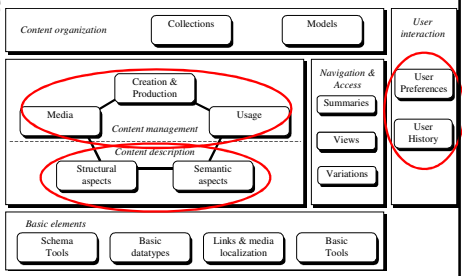


W3C

98

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



Part 5 - MDS

99

Image Description Issue: Describing parts of an image

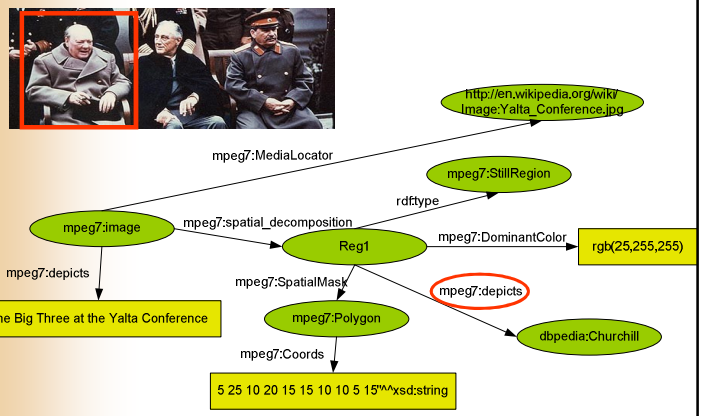
The "Big Three" at the Yalta Conference (Wikipedia)

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

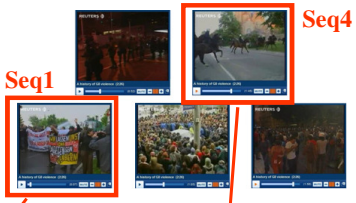
```
:Reg1 foaf:depicts dbpedia:WinstonChurchill
dbpedia:Churchill rdfs:label "Winston Churchill"
dbpedia:Churchill rdf:type foaf:Person
```

100

Image Description using MPEG-7



Video Description Issue: Describing parts of a video

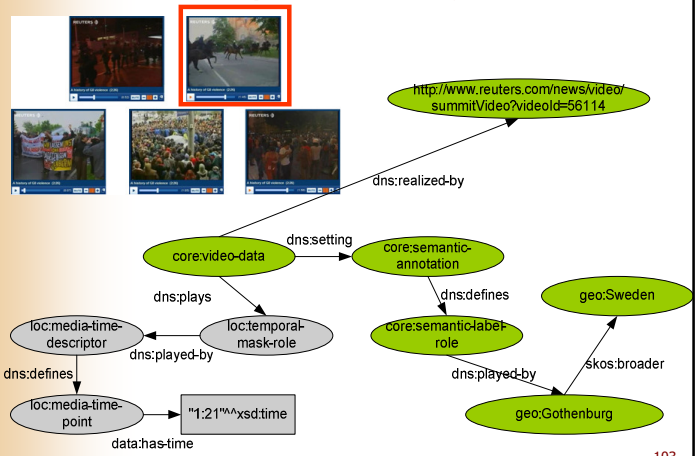


A history of G8 violence (video) (© Reuters)

- Localize a segment
- Annotate the content
  G8 Summit, Heiligendamm, 2007
  EU Summit, Gothenburg, 2001
- Link to the knowledge on the Web

```
:Seq1 foaf:depicts wikipedia:34th_G8_Summit
:Seq4 foaf:depicts wikipedia:EU_Summit
geo:Heiligendamm skos:broader geo:Germany
```


Video Description using MPEG-7



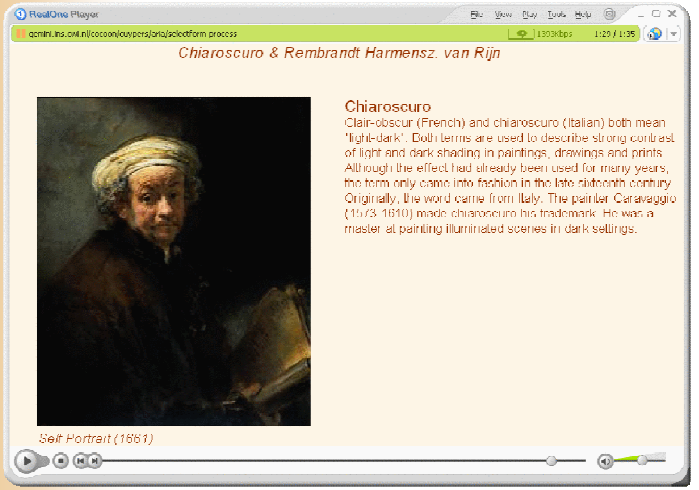
Agenda

1. Image/Video: search and description problems
2. A Multimedia Semantic Web:
   - Introduction to RDF, RDFS and OWL
   - The Linked Data Principle
   - The return of MPEG-7
3. Generate Multimedia Presentations

The problem

- Too many users need their own information
  - for their level of expertise
  - using appropriate media
  - in an appropriate style
  - displayed on their own device
- Multimedia information design is expensive
- There has to be some automation in the process

105



106

Three ingredients

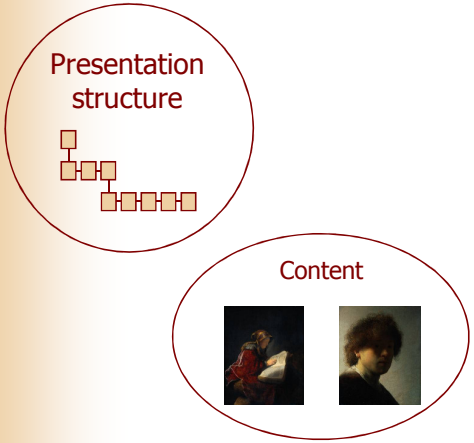
107

Content of example

Clair-obscur (Frans) en chiaroscuro (Italiaans) betekenen 'licht-donker'. Beide termen worden gebruikt om sterke licht-donkercontrasten in schilderijen, tekeningen en prenten aan te duiden. Hoewel het effect al eerder werd toegepast, is de term pas sinds het einde van de 16de eeuw in zwang. De oorsprong van het woord ligt in Italië. De schilder Caravaggio (1573-1610) maakte het chiaroscuro-effect tot zijn handelsmerk. Hij was een meester in het schilderen van donkere taferelen met één felle lichtbundel.

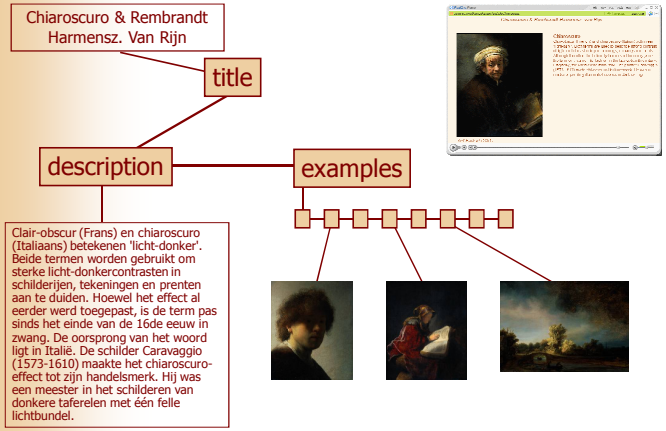
108

Three ingredients



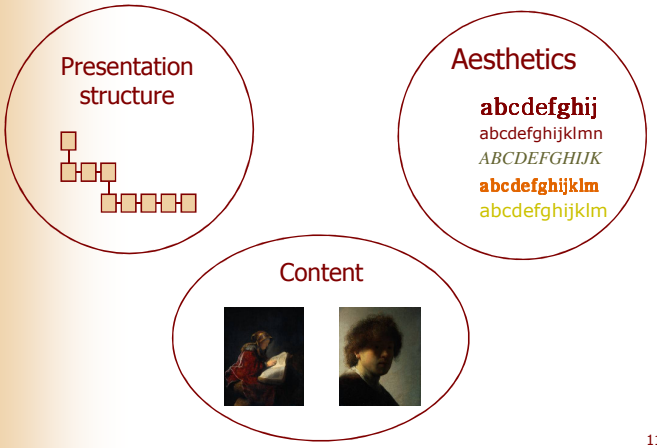
109

Presentation structure of example



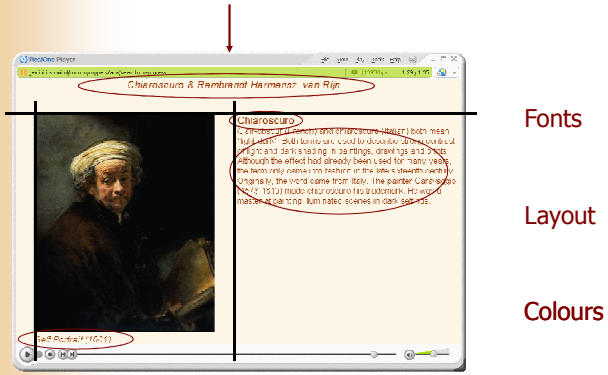
110

Three ingredients



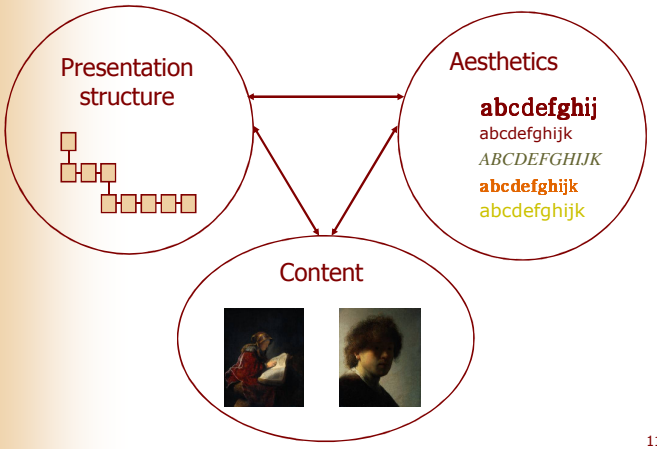
111

Aesthetics of example



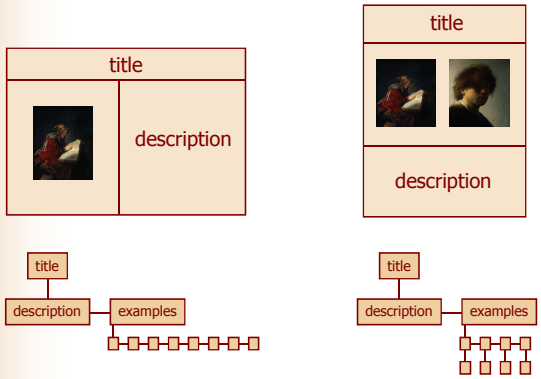
112

Design dependencies



113

Presentation Structure depends on Layout



114

Style Depends on Content

Chiaroscuro & Rembrandt

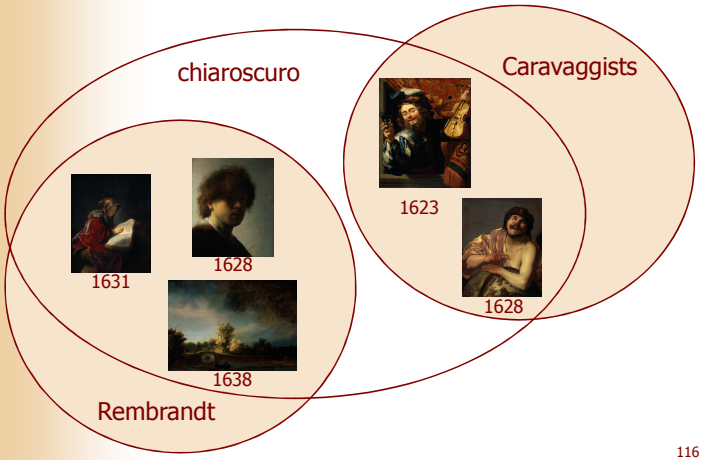
The Stone Bridge (1638)

Chiaroscuro & Rembrandt

The Stone Bridge (1638)

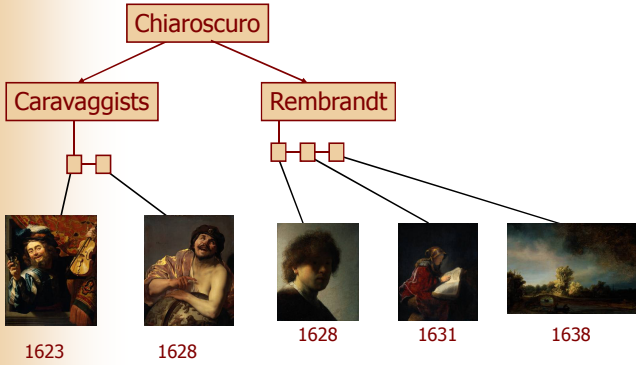
115

Presentation structure depends on content



116

Example Presentation Structure



117

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

119

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*

120

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

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

Scenarios

- Augmenting one interview with man-on-the-street opinion ("Vox Populi" documentary)
- Overview of the content of video footage:
  - Example: trailers ("Voices of Iraq")
  - Browse the content by opinion

123

The annotations

- Rhetorical
  - Rhetorical Statement (mostly verbal, but visual also possible)
  - Argumentation model: Toulmin model
- Descriptive
  - Question asked
  - Interviewee (social)
  - Filmic (e.g. location/time/framing/gaze)

124

Encode statements

- 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**

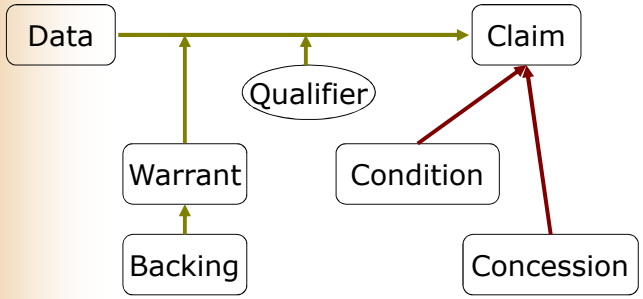
125

Connect statements

- Using the thesaurus, generate related statements and query the repository
  - E.g. from "*war best solution*" "*diplomacy best solution*", "*war not solution*"
- Create a graph of related statements
  - Nodes are the statements (video segments), edges are either *support* or *contradict*

126

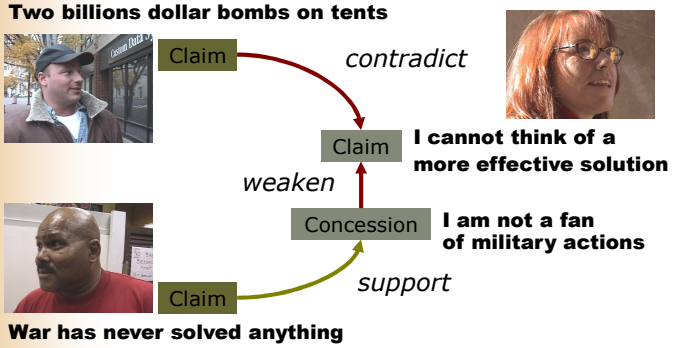
Toulmin model



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

127

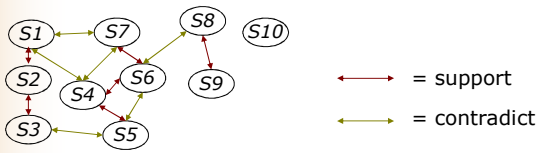
Analysis of the Example



128

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



129

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 and Against opinion X
- We use:
  - Logos (the statements)
  - Ethos (based on user profile)
  - Film editing (framing, gaze, counterpoint editing)

130

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

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

131