



Siderean™
SOFTWARE

Navigation for the Digital Universe

>
accenture

High performance. Delivered.

Semantic Web Techniques for Searching and Navigation Video Shots in BBC Rushes

Bradley P. Allen and Valery A. Petrushin

Outline

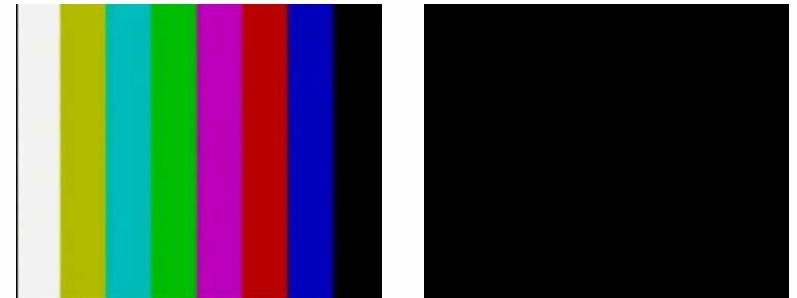
- Task
 - BBC Rushes exploration
- Approach
 - Build a faceted navigation interface exploiting both textual and visual metadata
- Implementation
 - BBC Rushes Navigator
 - Metadata representation
 - Architecture
 - User interface
 - Search + Fast Browsing System
- Additional Data Processing
- Recommendations

BBC Rushes 2006

- Rushes are raw footage ...
with a promise to turn into golden nuggets of stockshots
- BBC Rushes 2006
 - 96 tapes (or rolls) (51 h 41 m)
 - 48 tapes for development set (25 h 36 m)
 - 48 tapes for test set (26 h 05 m)
 - Canada (29 tapes – 15 h 19 m)
 - Senegal (32 tapes – 17 h 18 m)
 - Guadeloupe (35 tapes – 19 h 04 m)
 - All interviews are in French
- Metadata
 - No shot boundary segmentation
 - No keyframes
 - No keywords:
 - » Very poor textual description for Senegal and partially Guadeloupe tapes

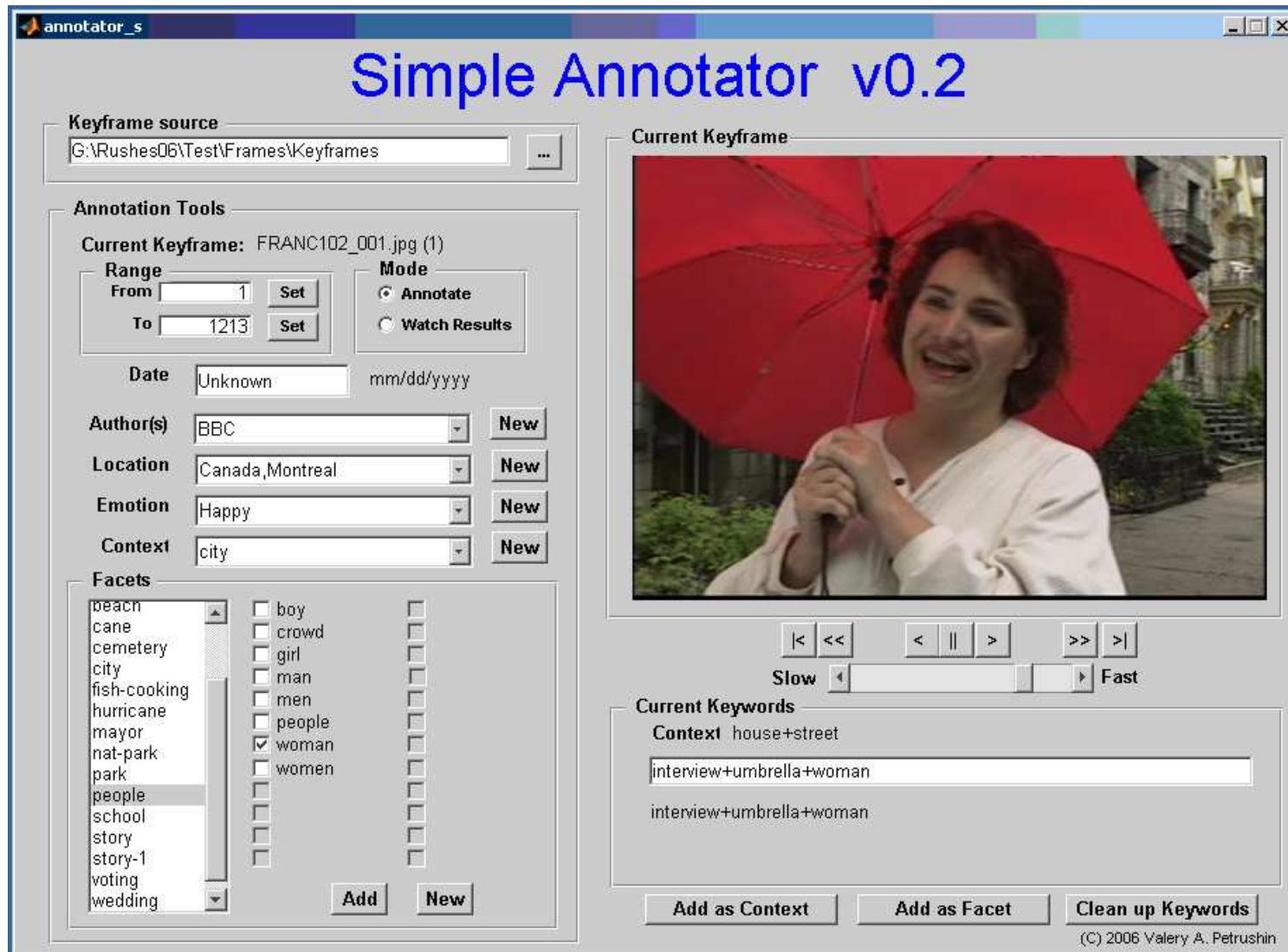
Data Preprocessing

- Shot boundary detection
 - Semi-automatic process
 - Automatic shot/keyframe extraction
 - Manual shot merging
- Shot selection
 - Deleting “junky” shots
 - Automatic recognition of “rainbow”, “black hole” and “grey field” shots
 - Selection short shots (transitions)
- Creating shot videos
 - Converting MPEG-1 into WMV
 - Splitting WMV files into shots
- Creating stories (clips)
 - Using a tool for manually assigning shot to stories (406 stories)
- Labeling keyframes
 - Using a tool for manually entering keywords

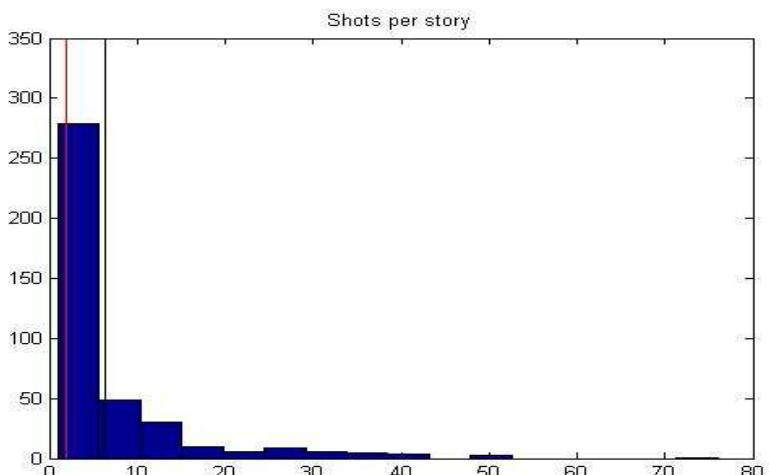
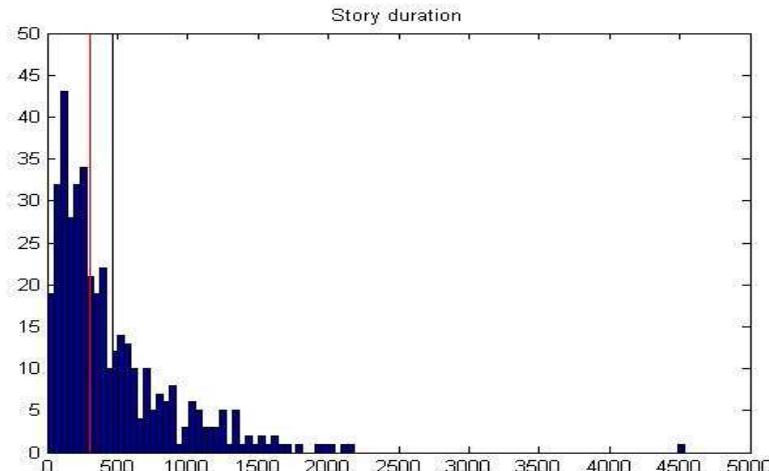




Snapshot of Annotation Tool



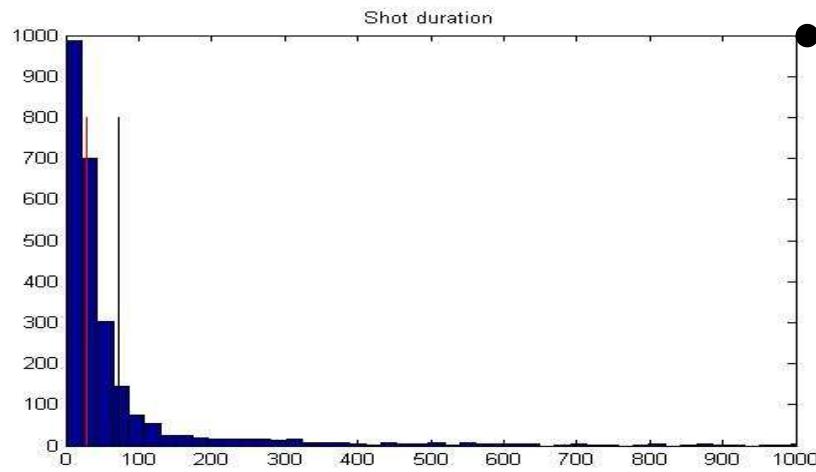
BBC Rushes 2006: data statistics - 1



- Statistics: story (clip) level

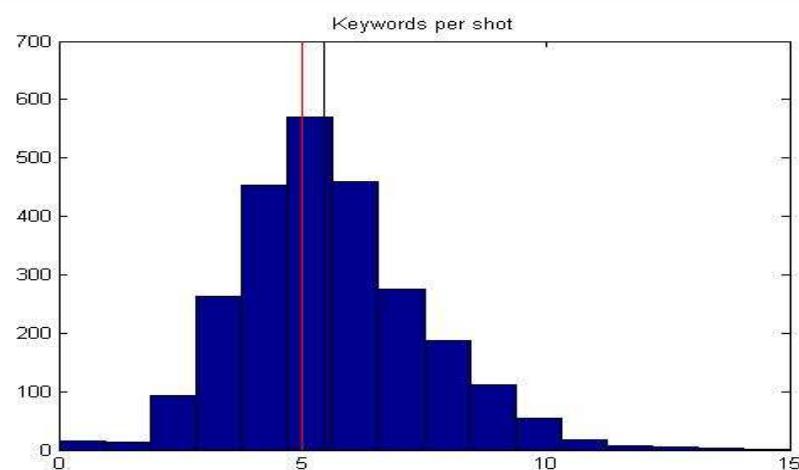
- Number of stories 406
- Number of interviews 201
- Story duration (hh:mm:ss)
 - » Minimal - 0:8.12
 - » Maximal – 1:15:35.72
 - » Mean – 7:42.62
 - » Median – 5:00.74
 - » Std - 7:41.11
- Number of shots 2550
- Number of shots per story
 - » Minimal - 1
 - » Maximal – 76
 - » Mean – 6.32
 - » Median – 2
 - » Std - 9.7

BBC Rushes 2006: data statistics - 2



- Statistics: shot level

- Number of shots 2,533
- Shot duration (mm:ss)
 - » Minimal - 0:00.16
 - » Maximal – 36:00.04
 - » Mean – 1:13.19
 - » Median – 0:27.92
 - » Std - 2:38.71



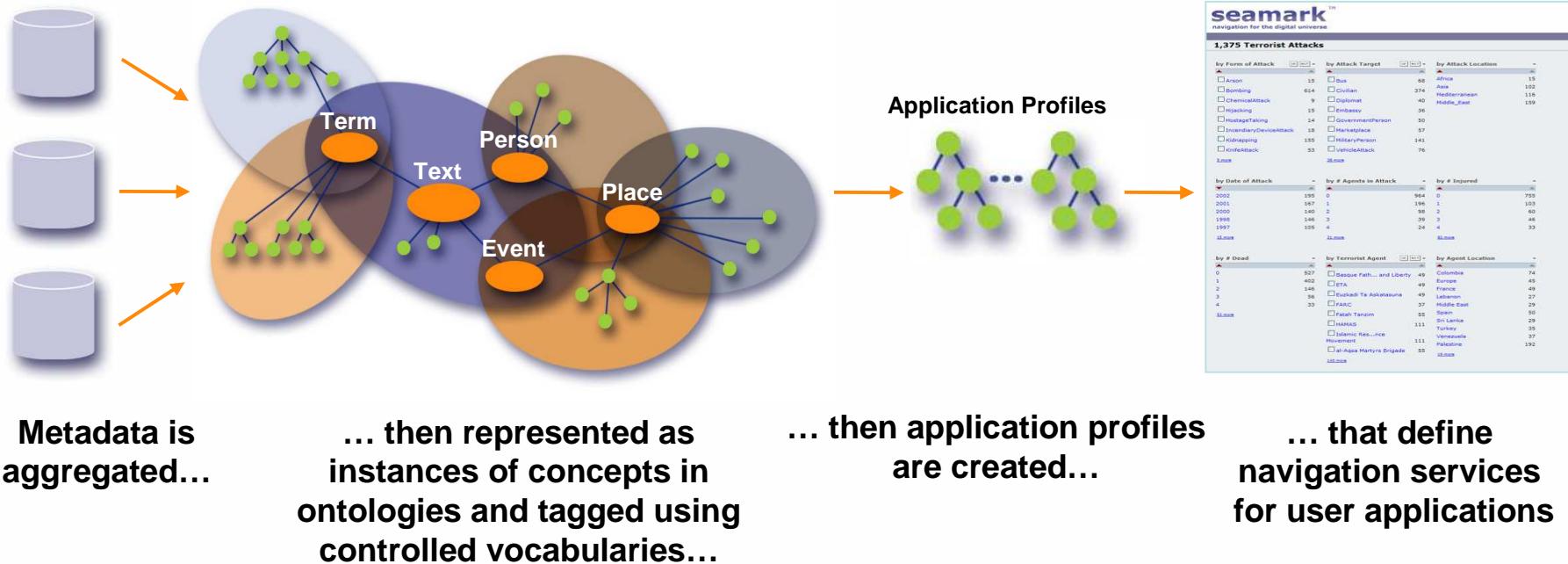
- Number of keywords

- » Total: 742
- » Occurrences: 13,801
- » Median per shot: 5
- » Mean per shot: 5.45
- » Maximal per shot : 15
- » Minimal per shot : 0
- » Std: 2.04

Faceted navigation

- *Facets* are metadata properties whose ranges form a near-orthogonal set of controlled vocabularies
 - Creator: “Dickens, Charles”
 - Subject: Arsenic, Antimony
 - Location: World > U.S. > California > Venice
- Facets form a frame of reference for information overview, access and discovery
 - Other properties serve as landmarks and cues
- *Faceted navigation* uses facets to provide end user access and discovery in the context of large collections of semi-structured information

Building faceted navigation applications



Faceted navigation built using SW standards

- Define/reuse ontologies expressed in RDF(S)/OWL
 - Classes for defining instances and controlled vocabularies
 - Properties for facets and additional asset metadata attributes
- Import/transform aggregated instance metadata into an RDF representation
 - Resources referred to via URIs
 - Content and controlled vocabularies
- Write application profiles in terms of RDF

Semantic Web Technology

- RDF(S) – Resource Description Framework (Schema)
- Dublin Core
- SKOS – Simple Knowledge Organization System
- TGM-I & II – Thesaurus for Graphic Materials
- LSCOM – Large Scale Concept Ontology for Multimedia
- SMIL – Synchronized Multimedia Integration Language
- MPEG-7 – Multimedia Content Description Interface

RDF (S)

- RDF (S) - Resource Description Framework (Schema)
 - <http://www.w3.org/RDF/>
 - <http://www.w3.org/TR/rdf-schema/>
 - language for representing metadata about Web resources
 - Triple : subject – predicate -- > object
 - Example:

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
           xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#">
  <contact:Person rdf:about="http://www.accenture.com/techlabs/VAP/contact#me">
    <contact:fullName>Valery A. Petrushin</contact:fullName>
    <contact:mailbox rdf:resource="mailto:valery.a.petrushin@accenture.com"/>
  </contact:Person>
</rdf:RDF>
```

Dublin Core (DC)

- Dublin Core

- <http://dublincore.org/documents/>
- vocabulary for describing documents (title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, relation, coverage, rights)
- Example:

```
<?xml version="1.0"?>
<!DOCTYPE rdf:RDF PUBLIC "-//DUBLIN CORE//DCMES DTD 2002/07/31//EN"
 "http://dublincore.org/documents/2002/07/31/dcmes-xml/dcmes-xml-dtd.dtd">
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
           xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://www.accenture/techlabs/Petrushin">
    <dc:title> Multimedia Data Mining and Knowledge Discovery</dc:title>
    <dc:creator> Valery A. Petrushin </dc:creator >
    <dc:publisher>Springer Verlag</dc:publisher>
  </rdf:Description>
</rdf:RDF>
```

SKOS

- SKOS – Simple Knowledge Organization System
 - <http://www.w3.org/2004/02/skos/>
 - model for expressing structure and content of concept schemes (thesauri, taxonomies, etc.)
 - Specifies concepts, collections of concepts and relations between concepts (broader, narrower, related)
 - Example:

```
<rdf:RDF  
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"  
    xmlns:skos="http://www.w3.org/2004/02/skos/core#">  
    <rdf:Description rdf:about="http://www.example.com/concepts#people">  
        <skos:broader rdf:resource="http://www.example.com/concepts#mammals"/>  
        <skos:narrower rdf:resource="http://www.example.com/concepts#children"/>  
        <skos:narrower rdf:resource="http://www.example.com/concepts#adults"/>  
    </rdf:Description>  
</rdf:RDF>
```

TGM – I & II

- TGM – Thesaurus for Graphic Materials (The Library of Congress)
 - TGM-I – Subject Terms (6,300)
 - <http://www.loc.gov/rr/print/tgm1/toc.html>
 - TGM-II – Genre and Physical Characteristic Headings (600)
 - <http://www.loc.gov/rr/print/tgm2/>
 - Example:

TGM-I:

Term: Sand

Narrower Term: Quicksand

Related Term: Dunes, Sand sculpture,
Sandpaintings

TGM-II:

Term: Aerial views

Public Note: Views from a high vantage point.

Used For: Air views, Balloon views, Views, Aerial

Broader Term: Views

Narrower Term: Aerial photographs

Related Term: Bird's-eye views, Panoramic views

LSCOM, SMIL & MPEG-7

- LSCOM – Large Scale Concept Ontology for Multimedia
 - http://www.acemedia.org/aceMedia/files/multimedia_ontology/presentations_1st_meeting/arda.pdf
- SMIL – Synchronized Multimedia Integration Language
 - <http://www.w3.org/TR/REC-smil/>
 - Simple language for representing multiple synchronized media streams
- MPEG-7 – Multimedia Content Description Interface
 - <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>
 - Advanced language for representing multimedia content
 - ISO Standard

BBC Rushes: representation

- Ontologies
 - RDFS, Dublin Core, SKOS
- Controlled vocabularies
 - TGM-1 (reflecting Light Scale Concept Ontology for Multimedia)
- Instances
 - trecvid:Shot, trecvid:Story
- Application profile
 - Retrieve instances of type trecvid:Shot
 - Textual facets: dc:title (clip title), dc:subject (keywords), dc:extent (duration), trecvid:emotion,
 - Motion facets: trecvid:intensity, trecvid:direction
 - Visual facets: dc:subject with values skos:narrower than trecvid:color, trecvid:texture and trecvid:shape

Self-organizing Maps

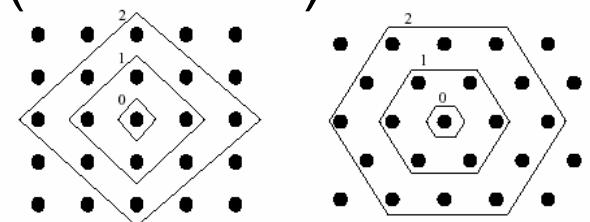
- SOM = Kohonen NN = Topology-preserving map
- Unsupervised learning (Clustering + Visualization)
- $X = \{x_i\}$, $x_i \in R^d$ - input data
- $M = \{m_k\}$, $m_k \in R^d$ - prototype vectors (codebook) = neurons on 1D or 2D grid
- Training:
 - 1. Start with random m_k
 - 2. For x_i find best-matching unit (BMU) m_c

$$\|x - m_c\| = \min_k \|x - m_k\|$$
 - 3. Update prototype vectors in neighborhood

$$m_k(t+1) = m_k(t) + \alpha(t)h_{ck}(t)[x(t) - m_i(t)]$$

where $h_{ck}(t)$ is the neighborhood kernel

$$h_{ck}(t) = \exp(-d_{ck}^2 / 2\sigma_t^2), \quad d_{ck} = \|r_c - r_k\|, \quad \sigma_t \text{ is radius at time } t$$
- Two phases: rough and fine tuning

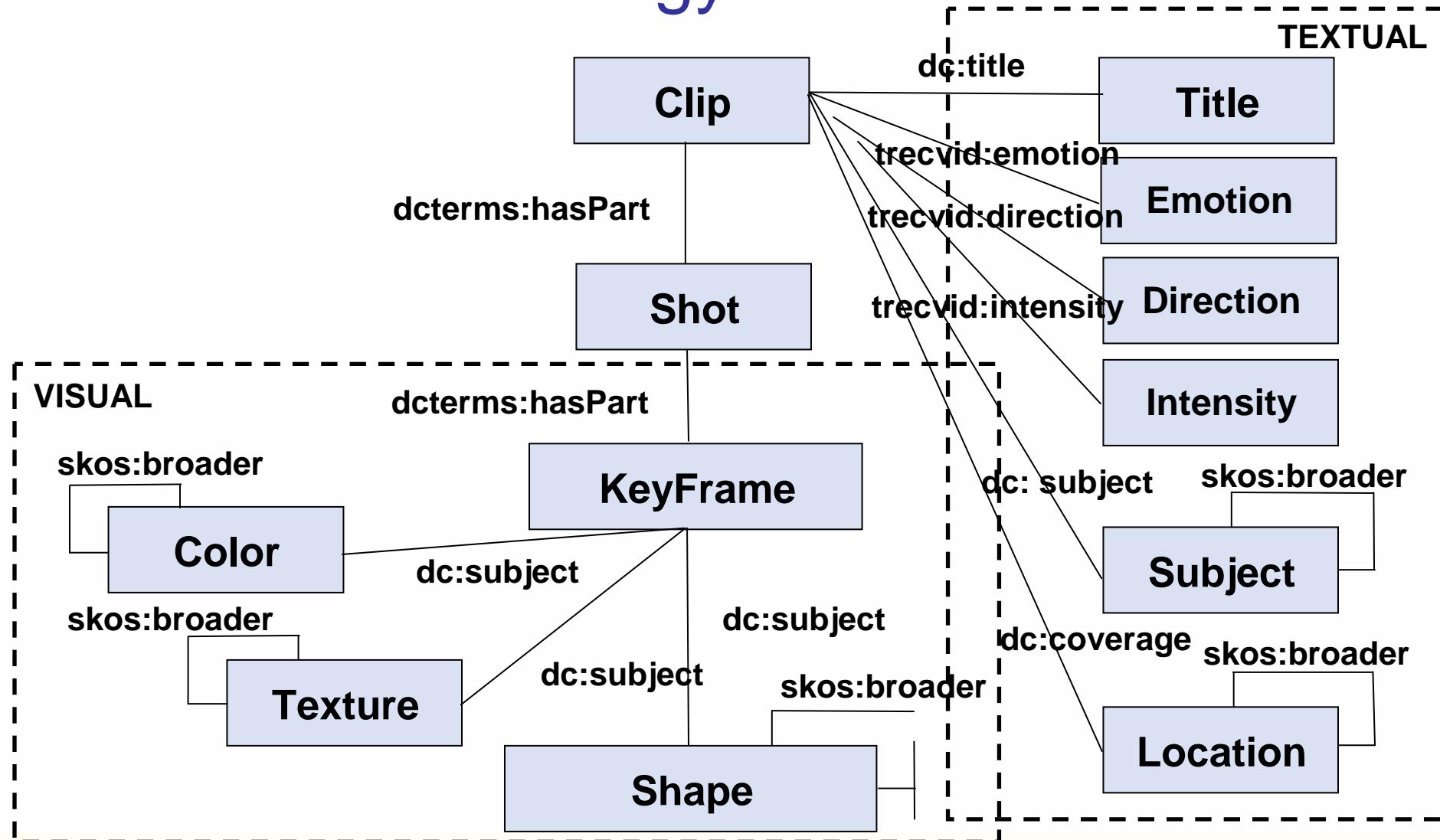




BBC Rushes: visual facets

- Facets: color, texture, shape
- To build facets
 - Extract features (MPEG-7):
 - Color: dominantColor, colorStructure, colorLayout, color histogram
 - Texture: edgeHistogram, homogenousTexture
 - Shape: contourShape, regionShape
 - SOM Clustering of keyframes
 - Select as a visual “word” the closest keyframe to node centroid
 - Represent keyframes as SKOS concepts, centroids as skos:broader of cluster members
- Example:
 - SOM for color 20 x 12

Ontology Schema





BBC Rushes: RDF/XML serialization

```
<trecvid:Shot rdf:about="http://swvideo.techlabs.accenture.com/Rushes2006/HTML/Shots/FRANC106_039.wmv">
  <rdf:type rdf:resource="&dtype;MovingImage" />
  <dcterms:extent>30680</dcterms:extent>
  <dc:relation>http://swvideo.techlabs.accenture.com/Rushes2006/HTML/Keyframes/FRANC106_039.jpg</dc:relation>
  <dcterms:hasPart rdf:resource="http://swvideo.techlabs.accenture.com/Rushes2006/HTML/Keyframes/FRANC106_039.jpg" />
  <trecvid:intensity>below average</trecvid:intensity>
  <trecvid:direction>upper left</trecvid:direction>
  <trecvid:emotion>neutral</trecvid:emotion>
  <dc:coverage>Canada</dc:coverage>
  <dc:coverage>Montreal</dc:coverage>
  <dc:subject rdf:resource="http://www.siderean.com/2005/08/trecvid/object#flag" />
  <dc:subject rdf:resource="http://www.siderean.com/2005/08/trecvid/object#monument" />
  <dc:subject rdf:resource="http://www.siderean.com/2005/08/trecvid/object#pedestal" />
  <dc:subject rdf:resource="http://www.siderean.com/2005/08/trecvid/object#sky" />
  <dc:subject rdf:resource="http://www.siderean.com/2005/08/trecvid/object#statue" />
  <dc:subject rdf:resource="http://www.siderean.com/2005/08/trecvid/object#tower" />
</trecvid:Shot>
<skos:Concept rdf:about="http://www.siderean.com/2005/08/trecvid/object#pedestal">
  <skos:prefLabel>pedestal</skos:prefLabel>
  <skos:broader rdf:resource="http://www.siderean.com/2005/08/trecvid#object" />
</skos:Concept>
<skos:Concept rdf:about="http://www.siderean.com/2005/08/trecvid/object#tower">
  <skos:prefLabel>tower</skos:prefLabel>
  <skos:broader rdf:resource="http://www.siderean.com/2005/08/trecvid#object" />
</skos:Concept>
```

BBC Rushes: RDF/XML serialization

```
<!-- color FRANC106_039 1 -->
<skos:Concept rdf:about="http://swvideo.techlabs.accenture.com/Rushes2006/HTML/Shots/color#FRANC106_039">
  <skos:prefLabel>FRANC106_039's color</skos:prefLabel>
  <skos:prefSymbol rdf:resource="http://swvideo.techlabs.accenture.com/Rushes2006/HTML/Shots/FRANC106_039.jpg" />
  <skos:broader rdf:resource="#trecvid;color" />
</skos:Concept>
<rdf:Description rdf:about="http://swvideo.techlabs.accenture.com/Rushes2006/HTML/Keyframes/FRANC105_047.jpg">
  <dc:subject rdf:resource="http://swvideo.techlabs.accenture.com/Rushes2006/HTML/Shots/color#FRANC106_039" />
</rdf:Description>
<rdf:Description rdf:about="http://swvideo.techlabs.accenture.com/Rushes2006/HTML/Keyframes/FRANC106_039.jpg">
  <dc:subject rdf:resource="http://swvideo.techlabs.accenture.com/Rushes2006/HTML/Shots/color#FRANC106_039" />
</rdf:Description>
<rdf:Description rdf:about="http://swvideo.techlabs.accenture.com/Rushes2006/HTML/Keyframes/FRANC108_019.jpg">
  <dc:subject rdf:resource="http://swvideo.techlabs.accenture.com/Rushes2006/HTML/Shots/color#FRANC106_039" />
</rdf:Description>
<rdf:Description rdf:about="http://swvideo.techlabs.accenture.com/Rushes2006/HTML/Keyframes/FRANC117_014.jpg">
  <dc:subject rdf:resource="http://swvideo.techlabs.accenture.com/Rushes2006/HTML/Shots/color#FRANC106_039" />
</rdf:Description>
```

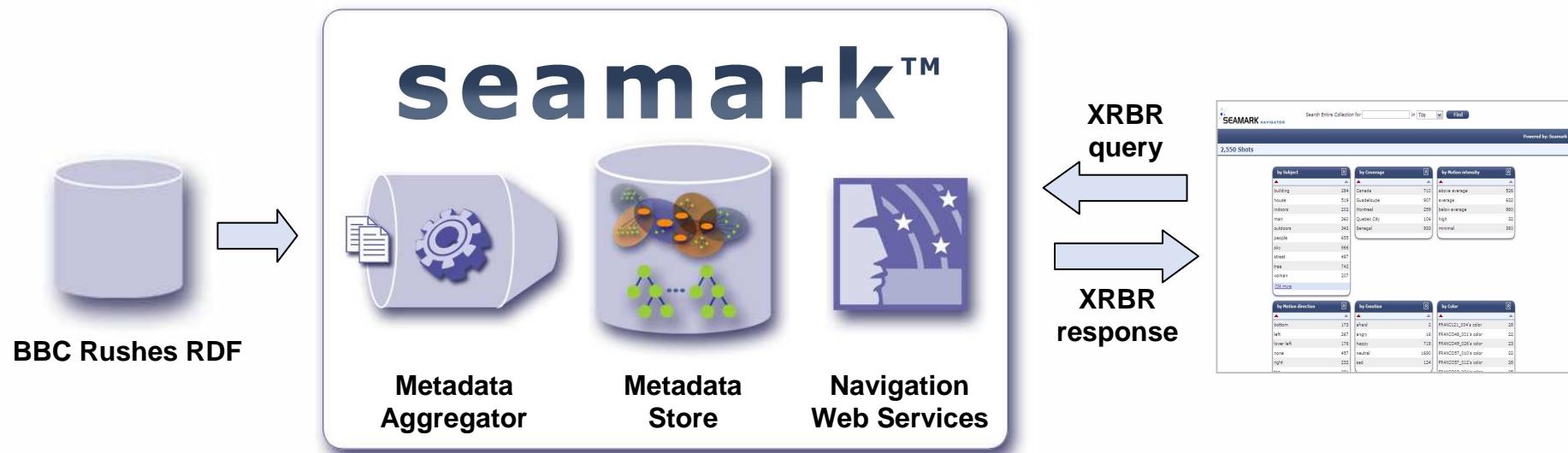


BBC Rushes: RDF/XML serialization

```
<trecvid:Story rdf:about="http://swvideo.techlabs.accenture.com/Rushes2006/Story#s0021">
  <rdf:type rdf:resource="&dctype;MovingImage" />
  <dc:title>Monuments</dc:title>
  <dcterms:extent>139040</dcterms:extent>
  <trecvid:emotion>Neutral</trecvid:emotion>
  <dc:coverage>Canada</dc:coverage>
  <dc:coverage>Montreal</dc:coverage>
  <dcterms:hasPart rdf:resource="http://swvideo.techlabs.accenture.com/Rushes2006/HTML/Shots/FRANC106_035.wmv" />
  <dcterms:hasPart rdf:resource="http://swvideo.techlabs.accenture.com/Rushes2006/HTML/Shots/FRANC106_036.wmv" />
  <dcterms:hasPart rdf:resource="http://swvideo.techlabs.accenture.com/Rushes2006/HTML/Shots/FRANC106_037.wmv" />
  <dcterms:hasPart rdf:resource="http://swvideo.techlabs.accenture.com/Rushes2006/HTML/Shots/FRANC106_038.wmv" />
  <dcterms:hasPart rdf:resource="http://swvideo.techlabs.accenture.com/Rushes2006/HTML/Shots/FRANC106_039.wmv" />
  <dcterms:hasPart rdf:resource="http://swvideo.techlabs.accenture.com/Rushes2006/HTML/Shots/FRANC106_040.wmv" />
  <dcterms:hasPart rdf:resource="http://swvideo.techlabs.accenture.com/Rushes2006/HTML/Shots/FRANC106_041.wmv" />
</trecvid:Story>

<rdf:Description rdf:about="http://swvideo.techlabs.accenture.com/Rushes2006/HTML/Shots/FRANC106_039.wmv">
  <dcterms:partOf rdf:resource="http://swvideo.techlabs.accenture.com/Rushes2006/Story#s0021" />
</rdf:Description>
```

BBC Rushes 2006 Navigator: architecture





BBC Rushes 2006 Navigator: user interface

A screenshot of the SEAMARK Navigator software interface. At the top, there's a search bar with the placeholder "Search Entire Collection for" and dropdown options for "in Title" or "in Description". Below the search bar is a "Find" button. To the right, it says "Powered by: Seemark". The main area displays six data tables under the heading "2,550 Shots".

- by Subject:**

building	284
house	519
indoors	202
man	360
outdoors	342
people	655
sky	666
street	487
tree	742
woman	207
734 more	
- by Coverage:**

Canada	710
Guadeloupe	907
Montreal	259
Quebec City	106
Senegal	933
- by Motion intensity:**

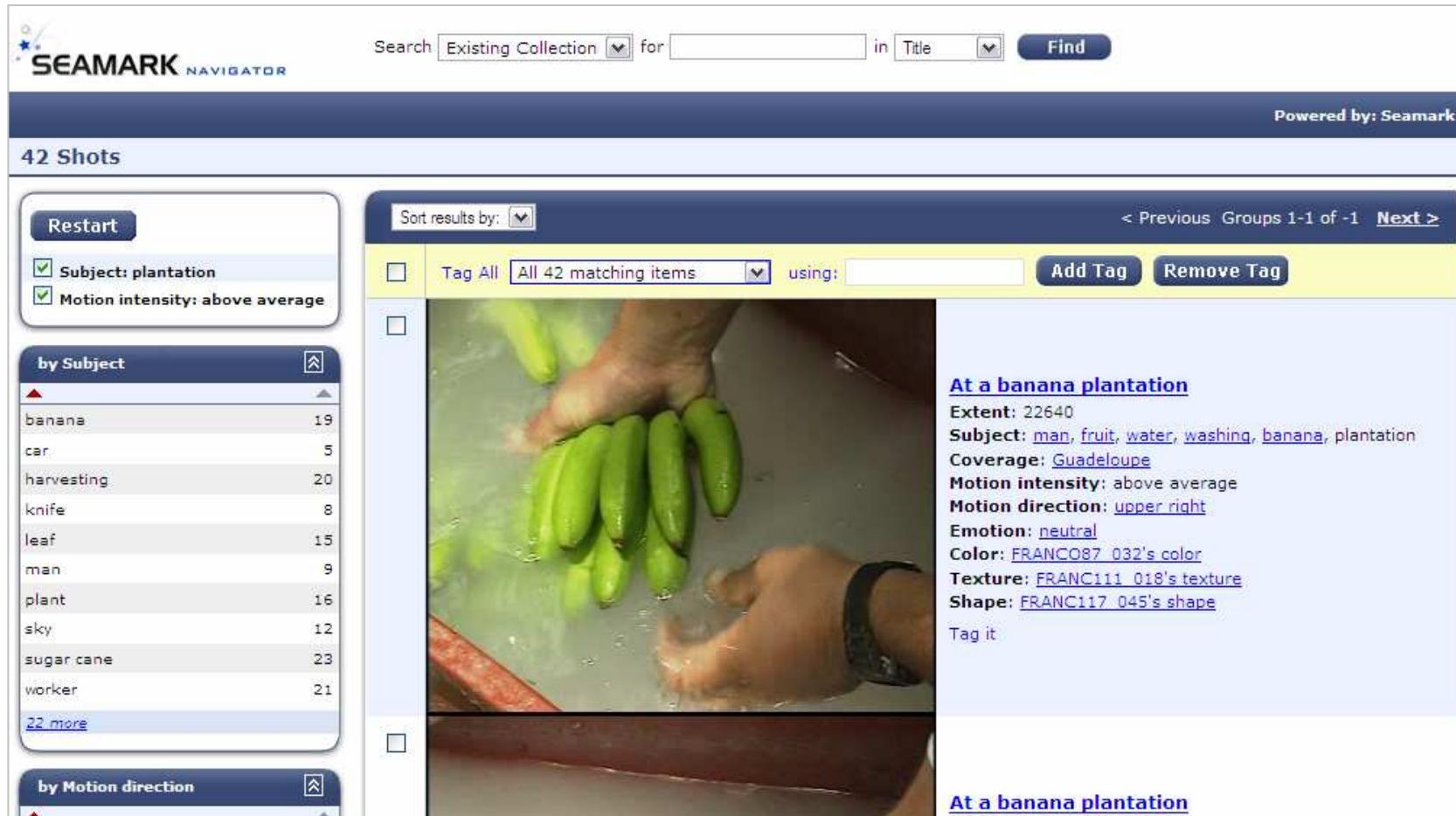
above average	526
average	632
below average	980
high	32
minimal	380
- by Motion direction:**

bottom	173
left	267
lower left	176
none	457
right	232
top	331
- by Emotion:**

afraid	2
angry	16
happy	718
neutral	1690
sad	124
- by Color:**

FRANC121_004's color	28
FRANCO48_001's color	22
FRANCO49_026's color	23
FRANCO57_010's color	22
FRANCO57_012's color	28
FRANCO68_001's color	25

BBC Rushes 2006 Navigator: user interface



The screenshot shows the SEAMARK NAVIGATOR interface. At the top, there is a search bar with the text "Search Existing Collection for [] in Title" and a "Find" button. To the right, it says "Powered by: Seemark". Below the search bar, the title "42 Shots" is displayed. On the left, there is a sidebar with a "Restart" button and two checked filters: "Subject: plantation" and "Motion intensity: above average". Below these filters is a list titled "by Subject" with items like banana (19), car (5), harvesting (20), knife (8), leaf (15), man (9), plant (16), sky (12), sugar cane (23), and worker (21). There is also a link to "22 more". At the bottom of this sidebar is another list titled "by Motion direction". The main area displays a grid of video thumbnails. One thumbnail is selected, showing a close-up of hands washing green bananas in water. To the right of this thumbnail, there is a detailed description of the shot:
At a banana plantation
Extent: 22640
Subject: [man](#), [fruit](#), [water](#), [washing](#), [banana](#), plantation
Coverage: [Guadeloupe](#)
Motion intensity: above average
Motion direction: upper right
Emotion: neutral
Color: [FRANCO87_032's color](#)
Texture: [FRANC111_018's texture](#)
Shape: [FRANC117_045's shape](#)
Tag it

Below this description, there is another link: **At a banana plantation**.

Additional Data Processing

- Interview segmentation into “question-response” chunks
 - Creating a speaker model for the interviewer (GMM-32 using MFCC)
 - Recognition the interviewer speech
 - Splitting interviews into Q-R chunks
- Automatic labeling using SOM
 - Create SOM for the development set
 - Use SOM as GMM classifier for keyframes from test set
 - Assign the frequent keywords derived from keyframes of the development set to keyframes from the test set
- Creating recognizers for semantic concepts using
 - Visual data
 - Water, sky, road, grass, building, etc.
 - Audio data
 - Music, speech, traffic noise, speaker’s gender, hammering, etc.



Search + Fast Browsing

The screenshot displays the Accenture Rapid Repository software interface. The top bar includes the application name, language (EN English (United States)), microphone icon, tools menu, and a help icon. The main window has a title bar "K Accenture Rapid Repository: Videos".

Search Criteria:

- Features: color (selected), texture, shape
- Motion intensity: Low (selected), High
- Motion direction: N, E, S, W
- Location: (dropdown menu)

Keywords For Search: interview_

Clip Information:

- Location: Guadeloupe
- Duration: 37.84"
- Emotion: Happy
- Motion Intensity: Low
- Motion direction: East

Keywords: boy grass house interview street tree

Favorites: (displaying small thumbnail images)

SEARCH RESULTS: A list of search results with file sizes in parentheses:

- CANADA (21 kB)
- CANADA (4 kB)
- CANADA, MONTREAL (6 kB)
- CANADA, MONTREAL (6 kB)
- CANADA, QUEBEC CITY (5 kB)
- SENEGAL (9 kB)
- SENEGAL (15 kB)
- SENEGAL (21 kB)
- SENEGAL (20 kB)
- GUADELOUPE (27 kB)
- GUADELOUPE (29 kB)
- GUADELOUPE (21 kB)
- GUADELOUPE (16 kB)
- GUADELOUPE (203 kB)

Thumbnail Preview: Shows a grid of 203 photos.

Recommendations

- Video database management
 - From shot extraction to semantic classification
 - Mining for stockshots
 - Interesting shots detection
 - New shots classification
 - Usage statistics
- Search for
 - Geographical location
 - Concrete concept (boat, fish, etc.)
 - Abstract concept (serenity, dance, etc.)
- Effectiveness and Usability
 - Interactive query model: fit of faceted navigation
 - Role of tagging in evaluation of exploration utility

Contact Information

- Bradley Allen

ballen@siderean.com

- Valery A. Petrushin

valery.a.petrushin@accenture.com

Thank you!