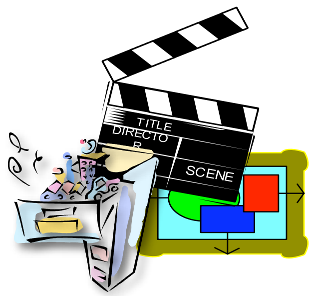
Ontologies and Machine Learning for Semantic Multimedia Analysis

Yiannis Kompatsiaris
Multimedia Knowledge Lab
CERTH
Informatics and Telematics Institute

Outline

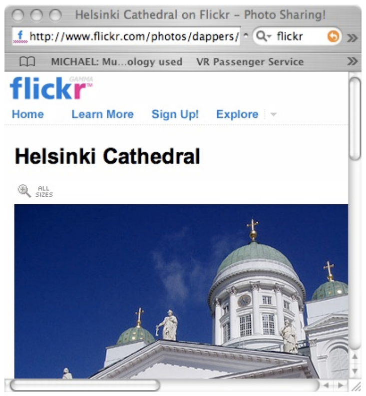
- Introduction
- Content - applications
- Semantic Multimedia Analysis
  - Multimedia Ontologies
  - Analysis
  - Reasoning
- Coclusions



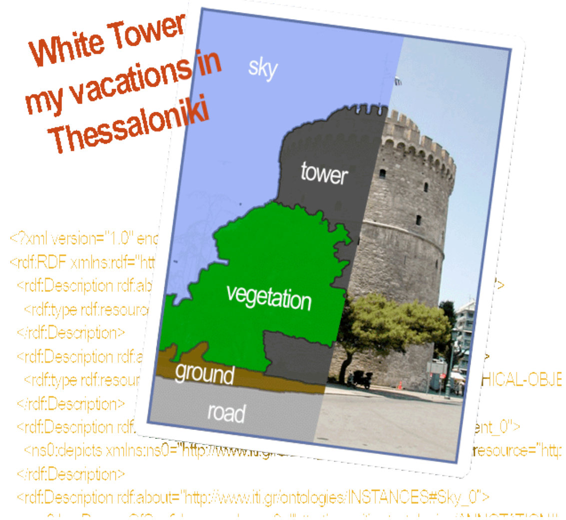
Multimedia Content

Networks

Storage & Devices



Web 2.0 photo - video applications



Segmentation KA Analysis

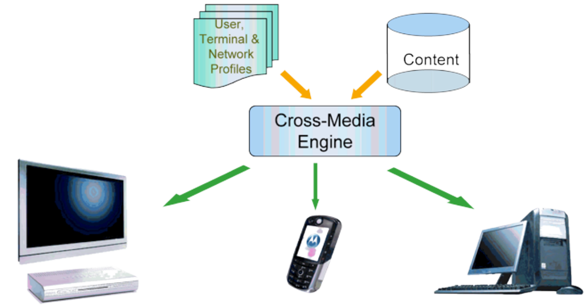
Labeling

Cross-media analysis

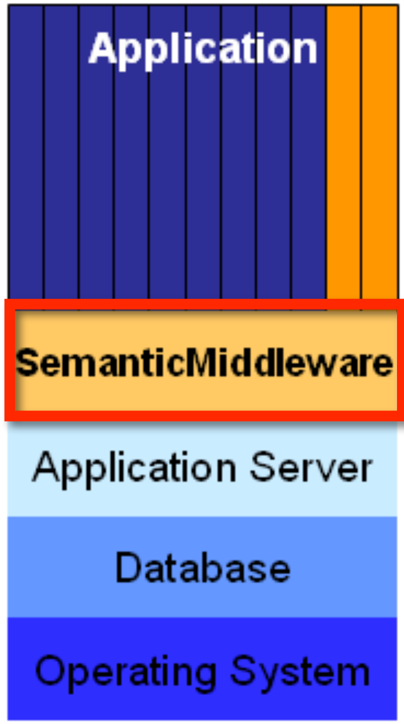
Context

Reasoning

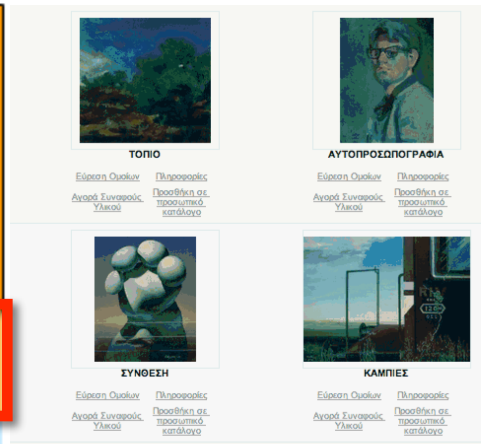
Metadata Generation & Representation



Content adaptation, personalization and distribution - Multiple Terminal & Networks



tomorrow



/ Content-based retrieval
dations and personalization

Semantic technology in Markets

Need for annotation

"The value of information depends on how easily it can be found, retrieved, accessed, filtered or managed in an active, personalized way"

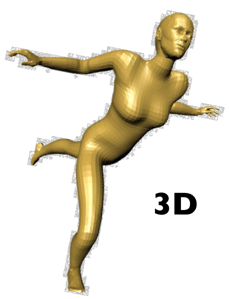
Content - Applications

Content → Knowledge Extraction → Applications

Personal

Sports

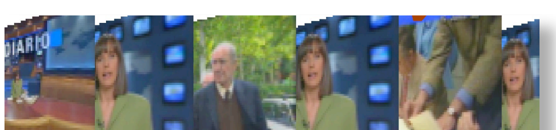
Fashion News

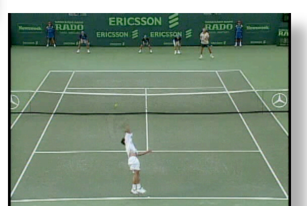


3D

Industrial

News





Commercial

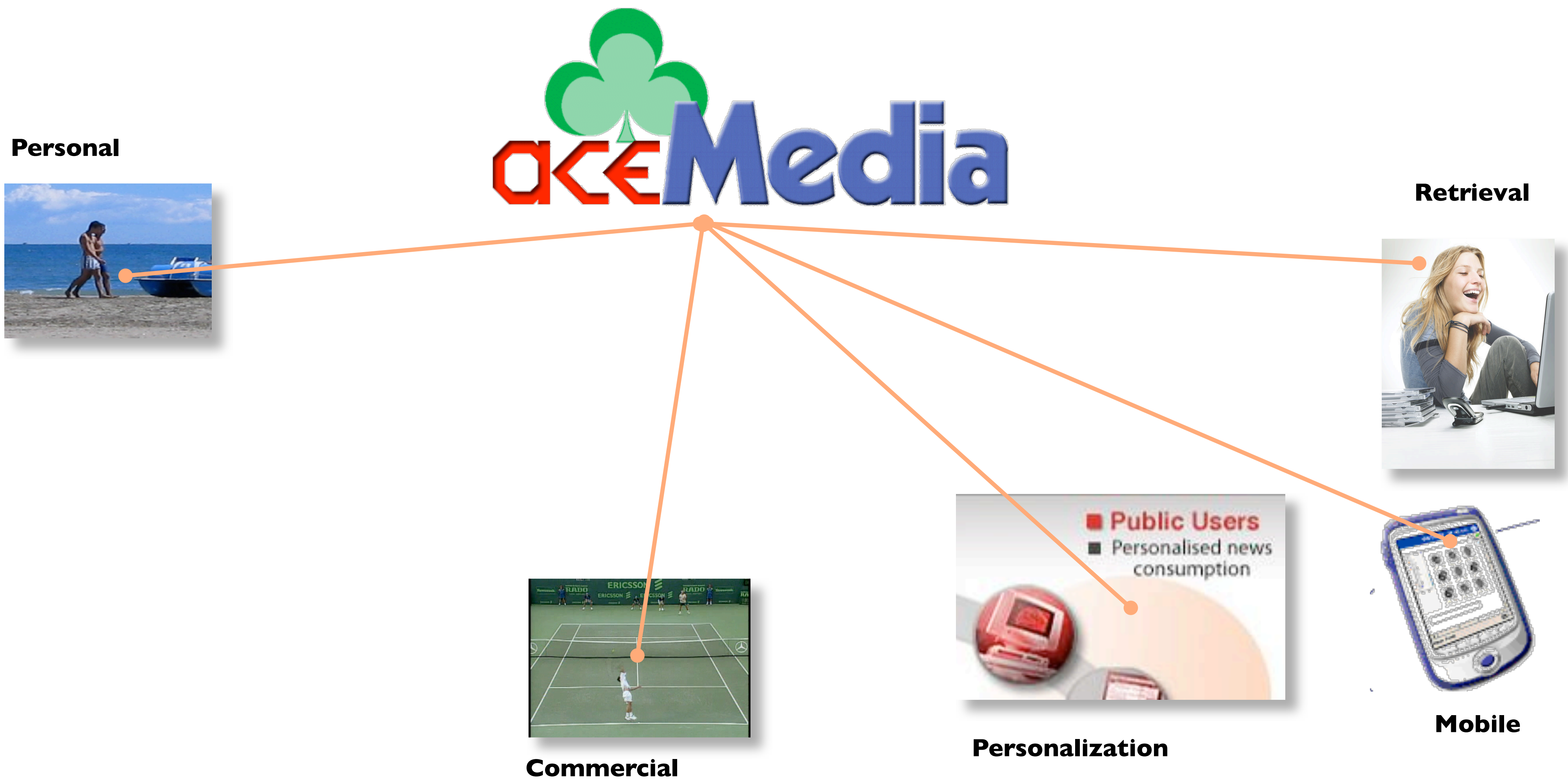
Semantic Desktop

Retrieval

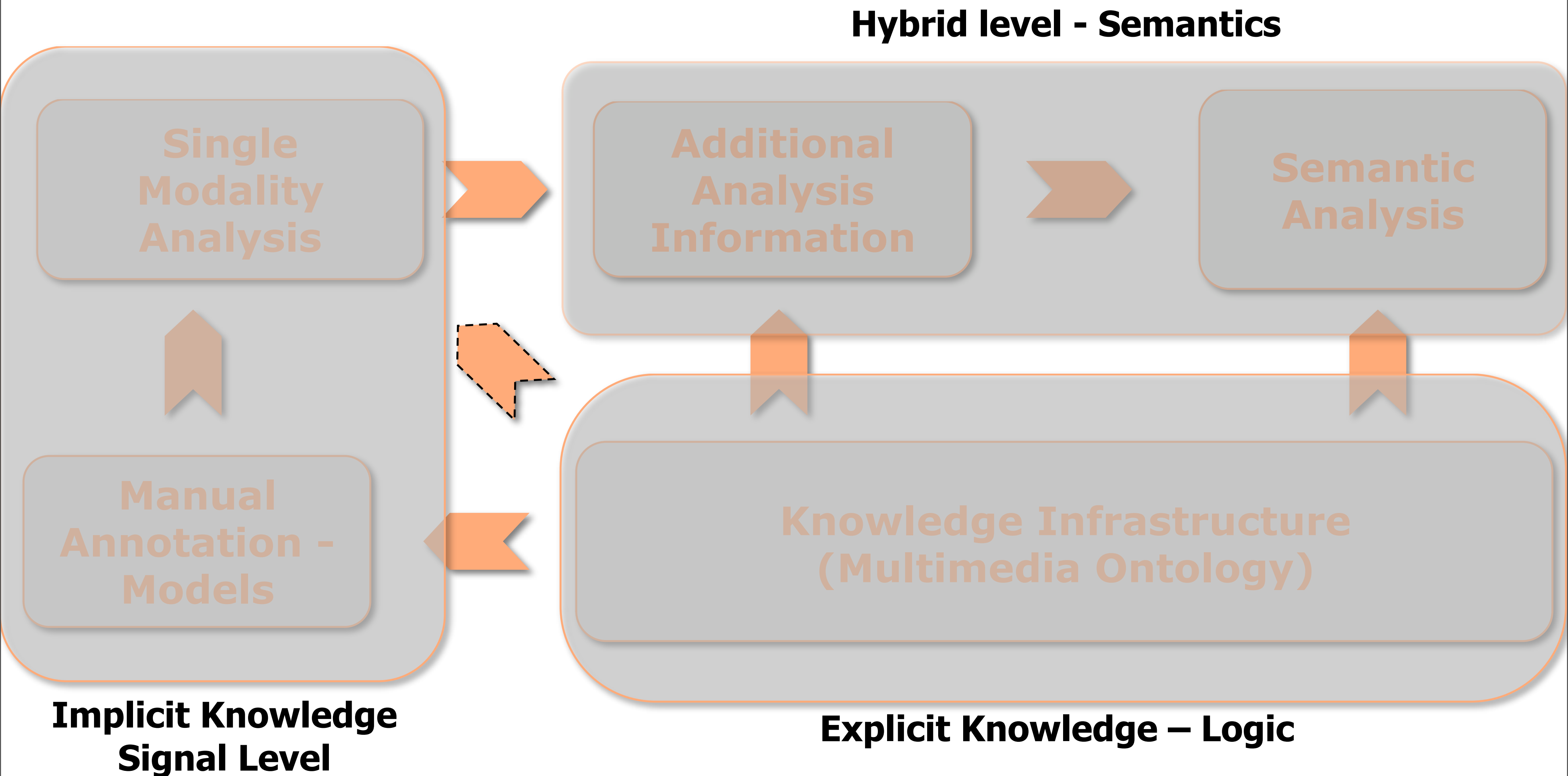
Personalization

Mobile

Content → Knowledge Extraction → Applications



Knowledge Extraction from MM



Knowledge Extraction from MM

LL feature extraction
Text, Image analysis
Segmentation, SVMs
<u>Evidence generation</u>
“Vehicle”, “Building”

Classifiers <u>fusion</u>
Global vs. Local
Modalities fusion
<u>Context</u>
“Ambulance”

<u>Reasoning</u>
Fusion of annotations and additional information
Consistency checking
“Emergency scene”

Single Modality Analysis

Additional Analysis Information

Semantic Analysis

Manual Annotation - Models

Knowledge Infrastructure (Multimedia Ontology)

Multimedia content <u>annotation</u> tools
Training
(Statistical) Modeling

Domain
Multimedia content
Annotations
Algorithms - Features
Context

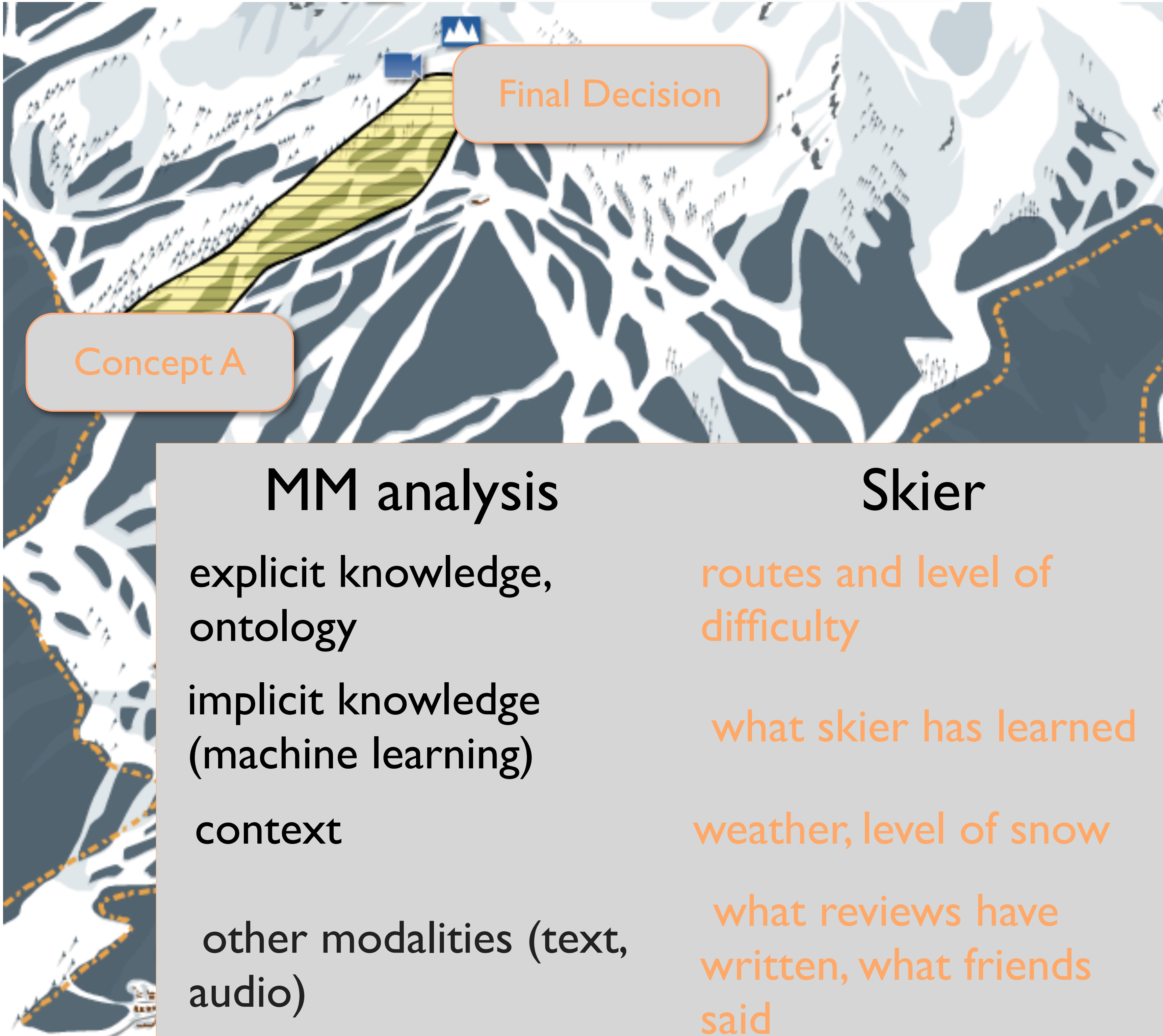
Addressing the *Semantic Gap*

Semantic Gap *for multimedia: To map automatically generated numerical low level-features to higher level human-understandable semantic concepts*


```
<?xml version='1.0' encoding='ISO-8859-1' ?>
<Mpeg7 xmlns…>
<DescriptionUnit xsi:type = "DescriptorCollectionType">
<Descriptor xsi:type = "DominantColorType">
<SpatialCoherency>31</SpatialCoherency>
<Value>
<Percentage>31</Percentage>
<Index>19  23  29 </Index>
<ColorVariance>0  0  0 </ColorVariance>
</Value>
</Descriptor>
</DescriptionUnit>
</Mpeg7>
```


This image contains a sky region and is a holiday image

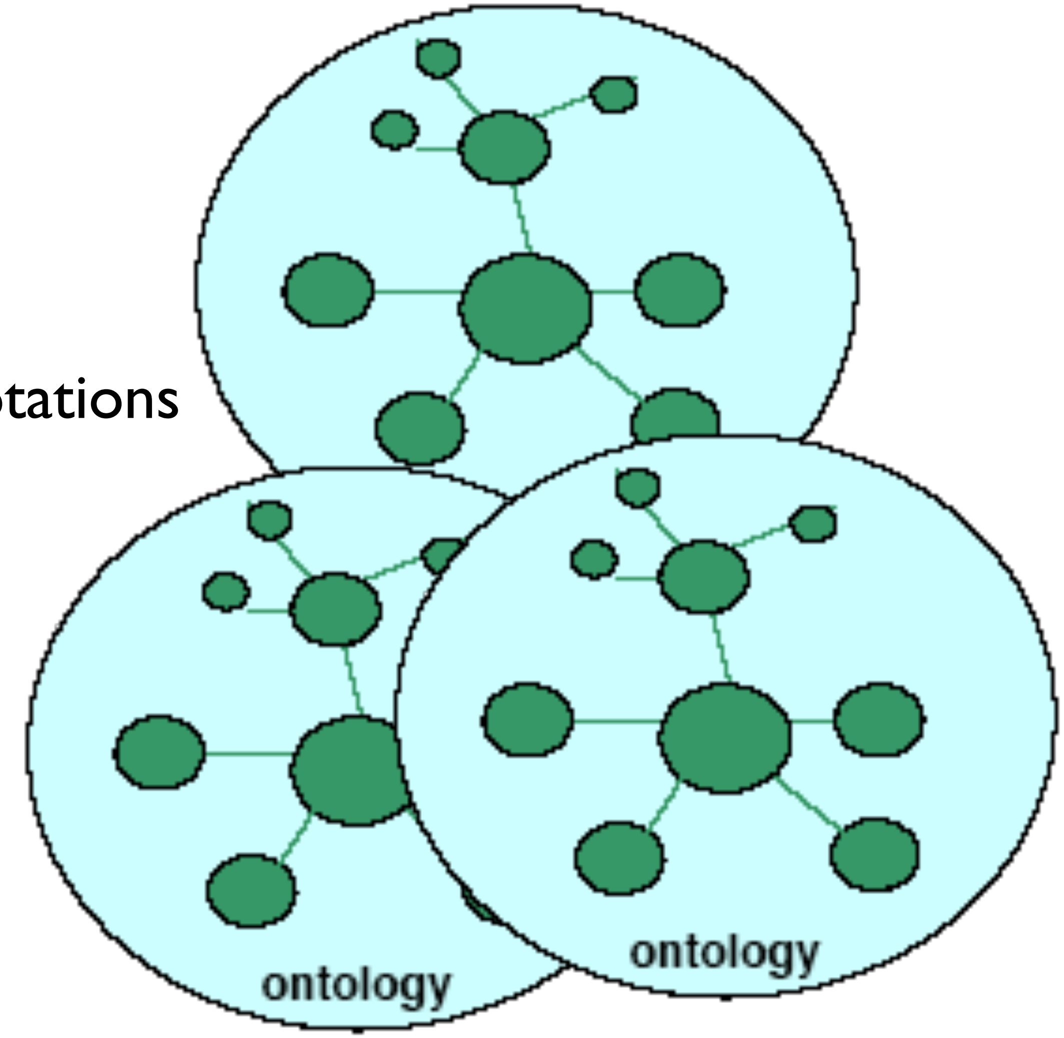
Dominant Color Descriptor of a sky region



MM analysis	Skier
explicit knowledge, ontology	routes and level of difficulty
implicit knowledge (machine learning)	what skier has learned
context	weather, level of snow
other modalities (text, audio)	what reviews have written, what friends said

Use of ontologies

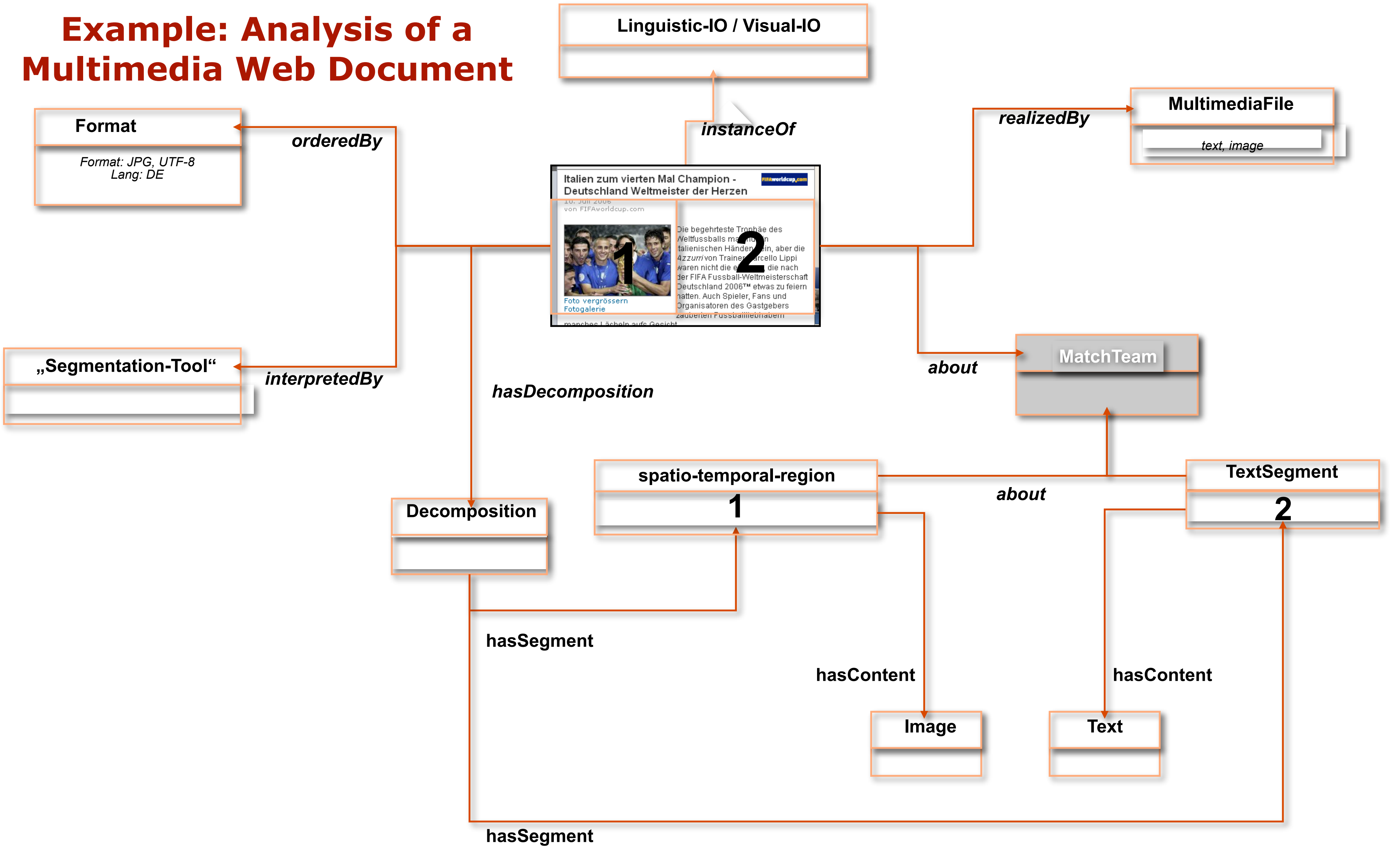
- Metadata representation
  - Annotation
  - Interoperability
- Reasoning
  - Extracting higher-level annotations
  - Consistency checking
  - Fusion
- Ontology-driven analysis
- Retrieval
- Personalization



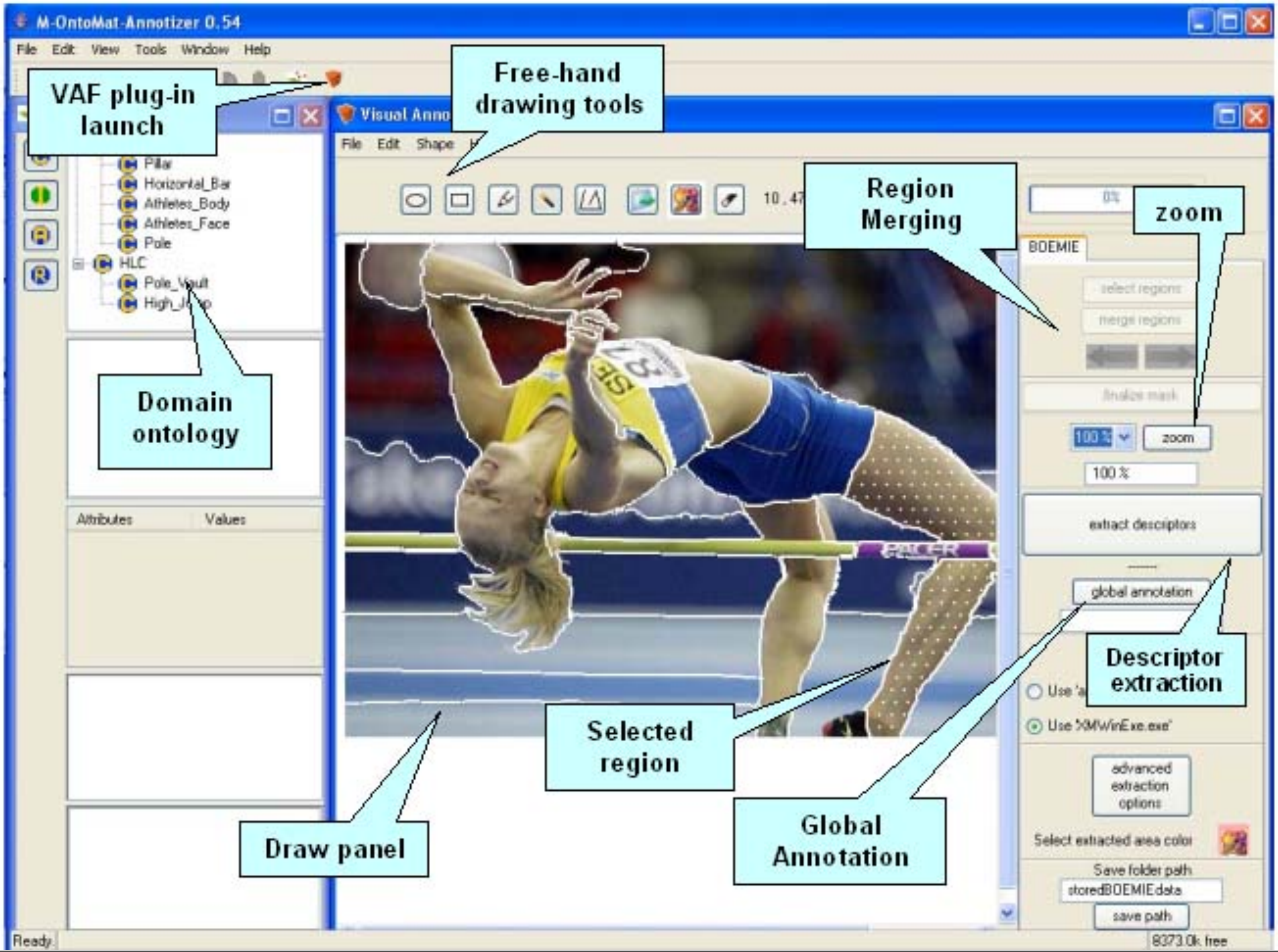
- Multimedia content structure
  - aceMedia (MPEG-7, RDF), AIM@SHAPE (3D content)
- Multimodality
  - MESH (OWL), BOEMIE (OWL-DL)
  - K-Space, X-Media (COMM, OWL, DOLCE)
- Fuzziness
  - K-Space, X-Media (Fuzzy-OWL)
- Changing knowledge
  - BOEMIE (evolution)
  - X-Media (versioning, reasons of change)
- Specific domains

Multimedia Information Objects -MESH

Example: Analysis of a Multimedia Web Document



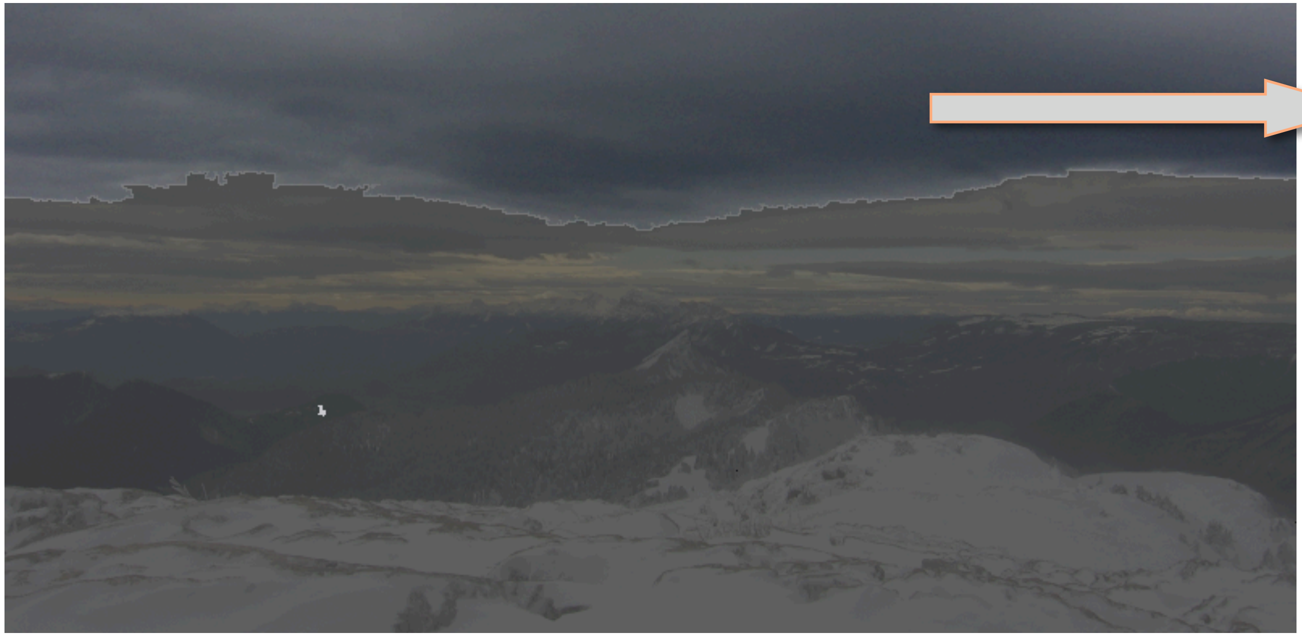
Annotation Tools



Multimedia Content Analysis

- MPEG-7 widely used for LL features
- Segmentation and feature extraction tools
- Well-known classifiers applied and developed
  - SVMs, EM, HMM - Bio-inspired approaches

Classifications results



Segment's hypothesis set

Natural-Person: 0.456798
Sailing-Boat: 0.463645
Sand: 0.476777
Building: 0.415358
Pavement: 0.454740
Road: 0.503242
Body-Of-Water: 0.489957
Cliff: 0.472907
Cloud: 0.757926
Mountain: 0.512597
Sea: 0.455338
Sky: 0.658825
Stone: 0.471733
Waterfall: 0.500000
Wave: 0.476669
Dried-Plant: 0.494825
Dried-Plant-Snowed: 0.476524
Foliage: 0.497562
Grass: 0.491781
Tree: 0.447355
Trunk: 0.493255
Snow: 0.467218
Sunset: 0.503164
Car: 0.456347
Ground: 0.454769
Lamp-Post: 0.499387
Statue: 0.501076

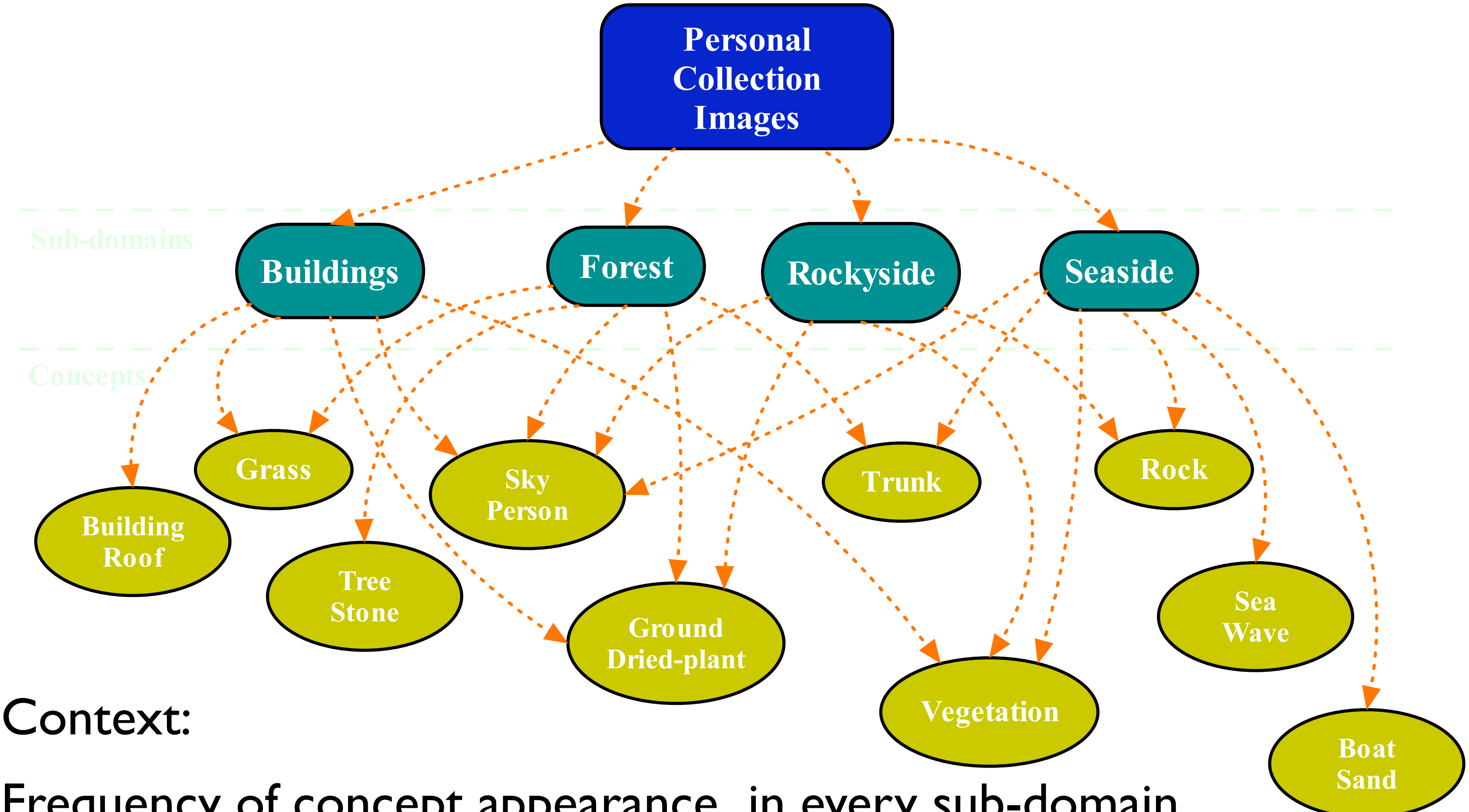
Multimedia Content Analysis

- Increasing use of context
  - Spatial, Frequency, EXIF (time, conditions, etc)
  - Recently: social context
- Fusion
  - Classifiers (global+local)
  - Modalities
    - e.g. Text+Image
    - Text+Speech+Video

Knowledge-Assisted Image Analysis
A specific approach

- Important observations:
  - Global-level information is not always sufficient
  - Local-level information can provide valuable cues
- The proposed approach combines:
  - Implicit and explicit knowledge
  - Global- and local-level information (Fusion)
  - Contextual information

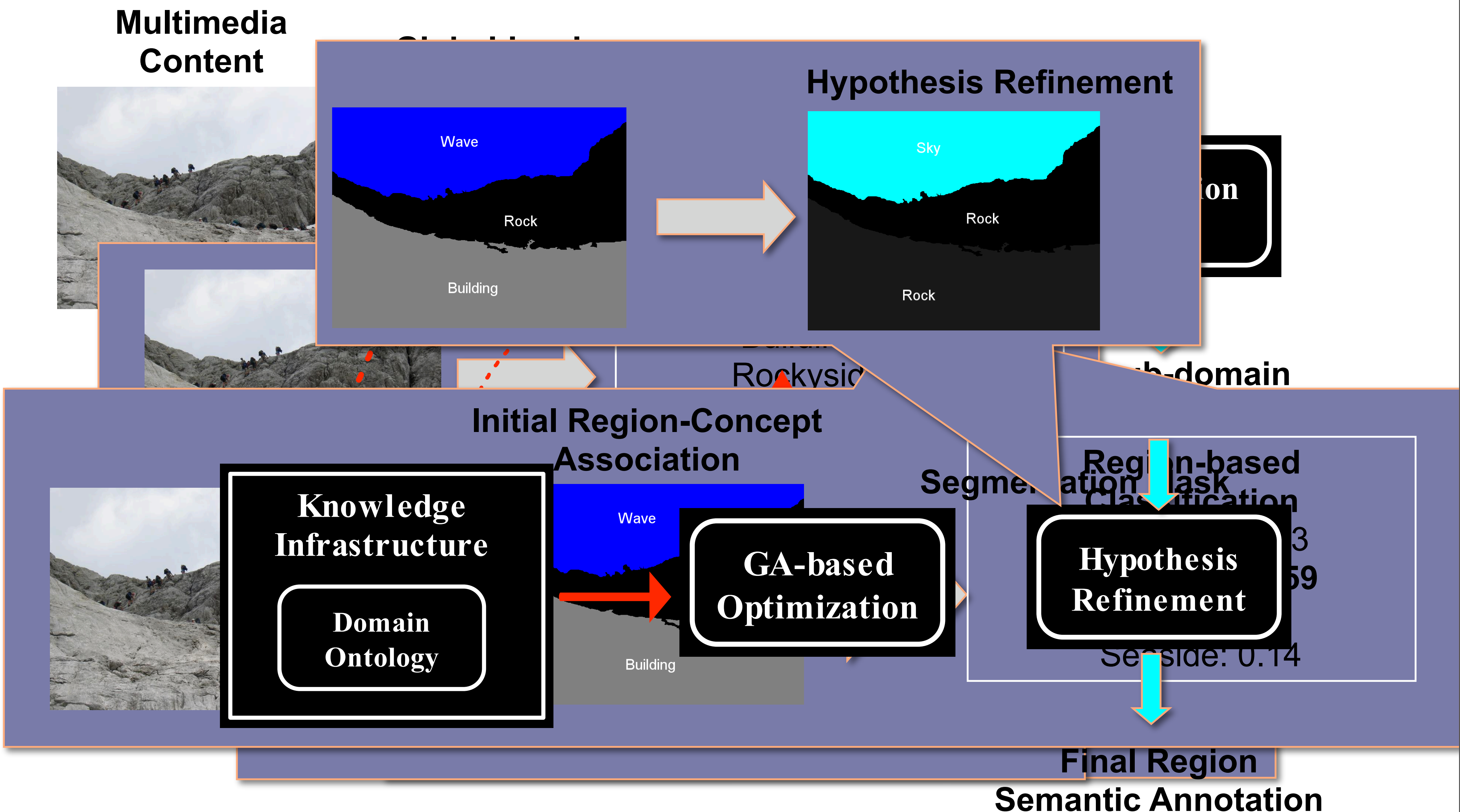
Knowledge Infrastructure



Context:

Frequency of concept appearance in every sub-domain

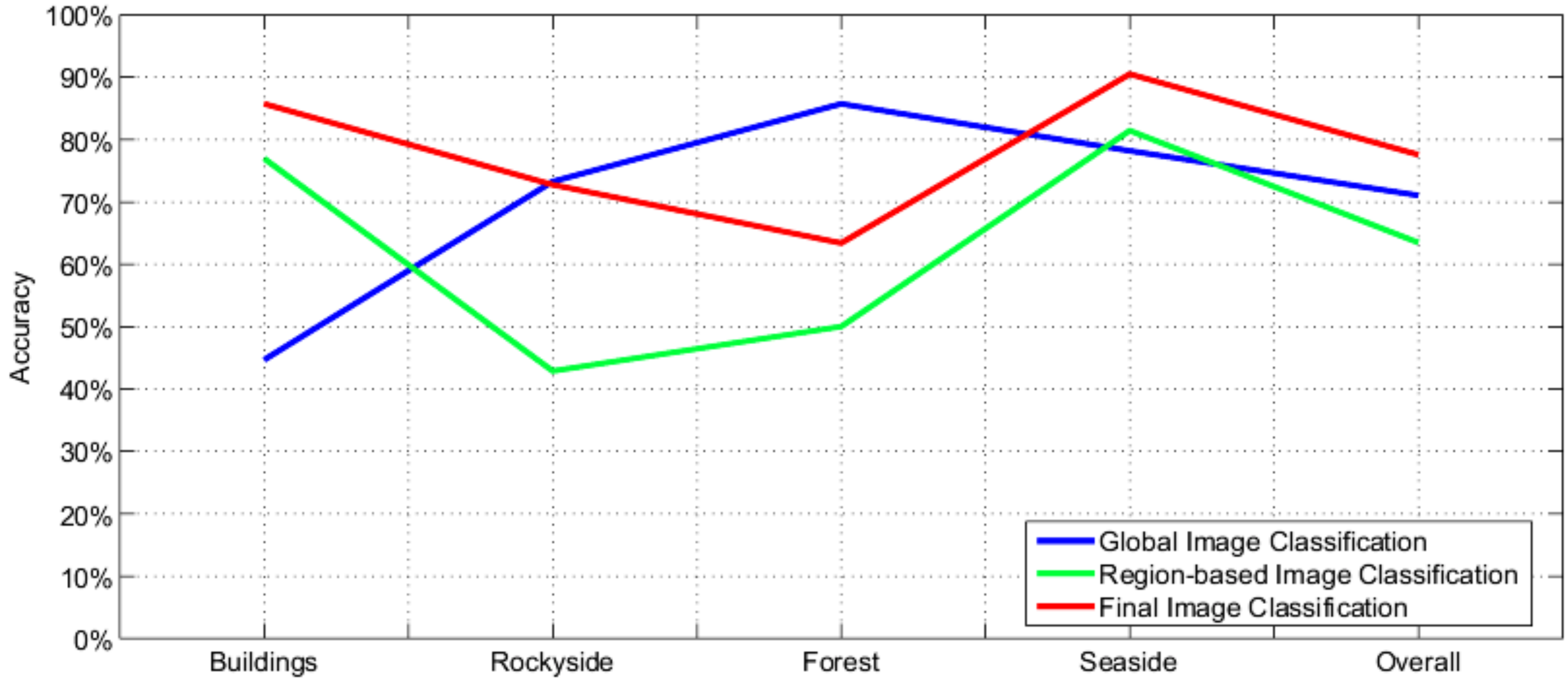
Proposed Approach



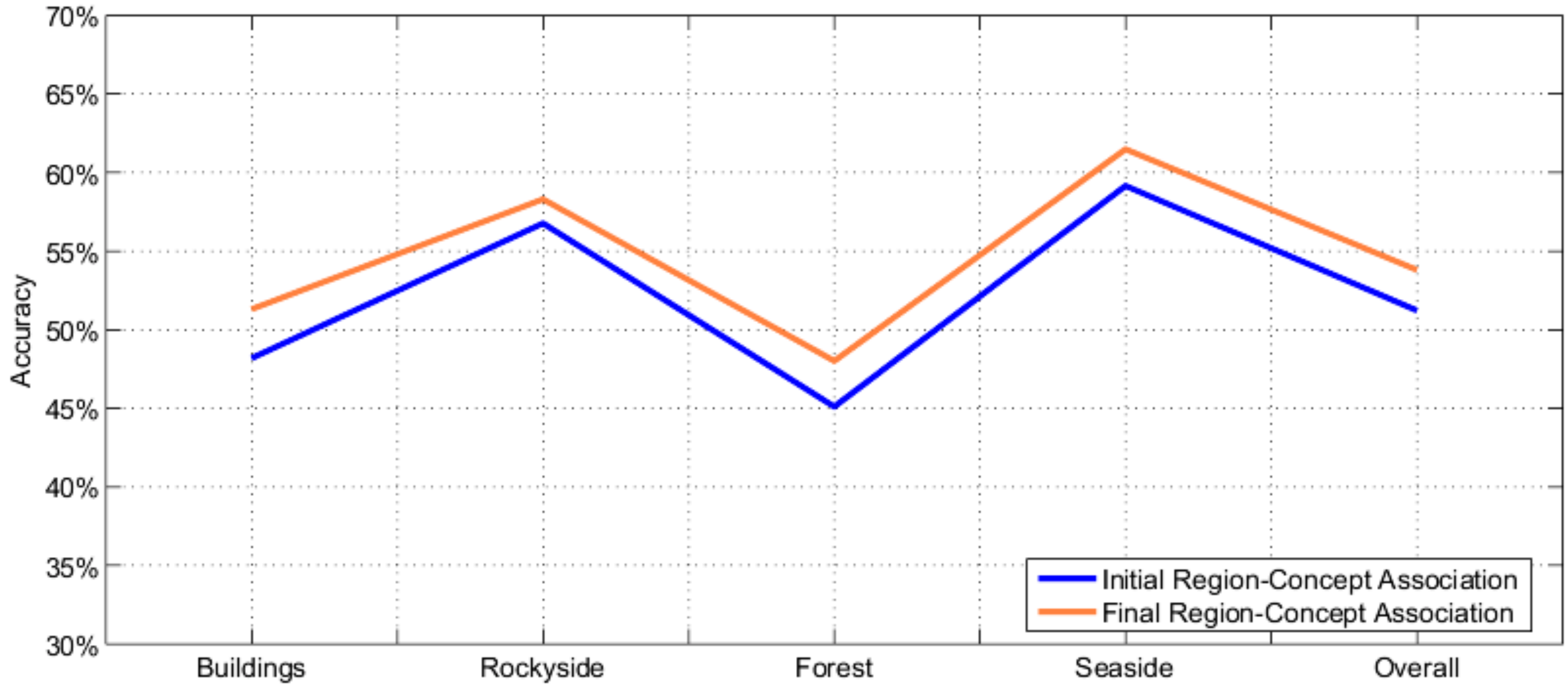
Experimental Results

- Domain of experimentation: Personal collection images domain
- 4 supported sub-domains
  - Buildings, Rockyside, Forest, Seaside
- 16 supported concepts
  - Building, Roof, Grass, Tree, Stone, Ground, Dried-plant, Sky, Person, Trunk, Vegetation, Rock, Boat, Sand, Sea, Wave
- 400 training images
- 400 testing images

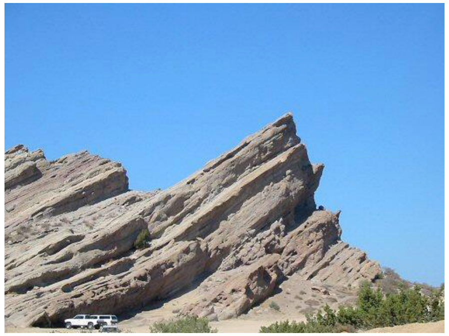
Sub-domain detection



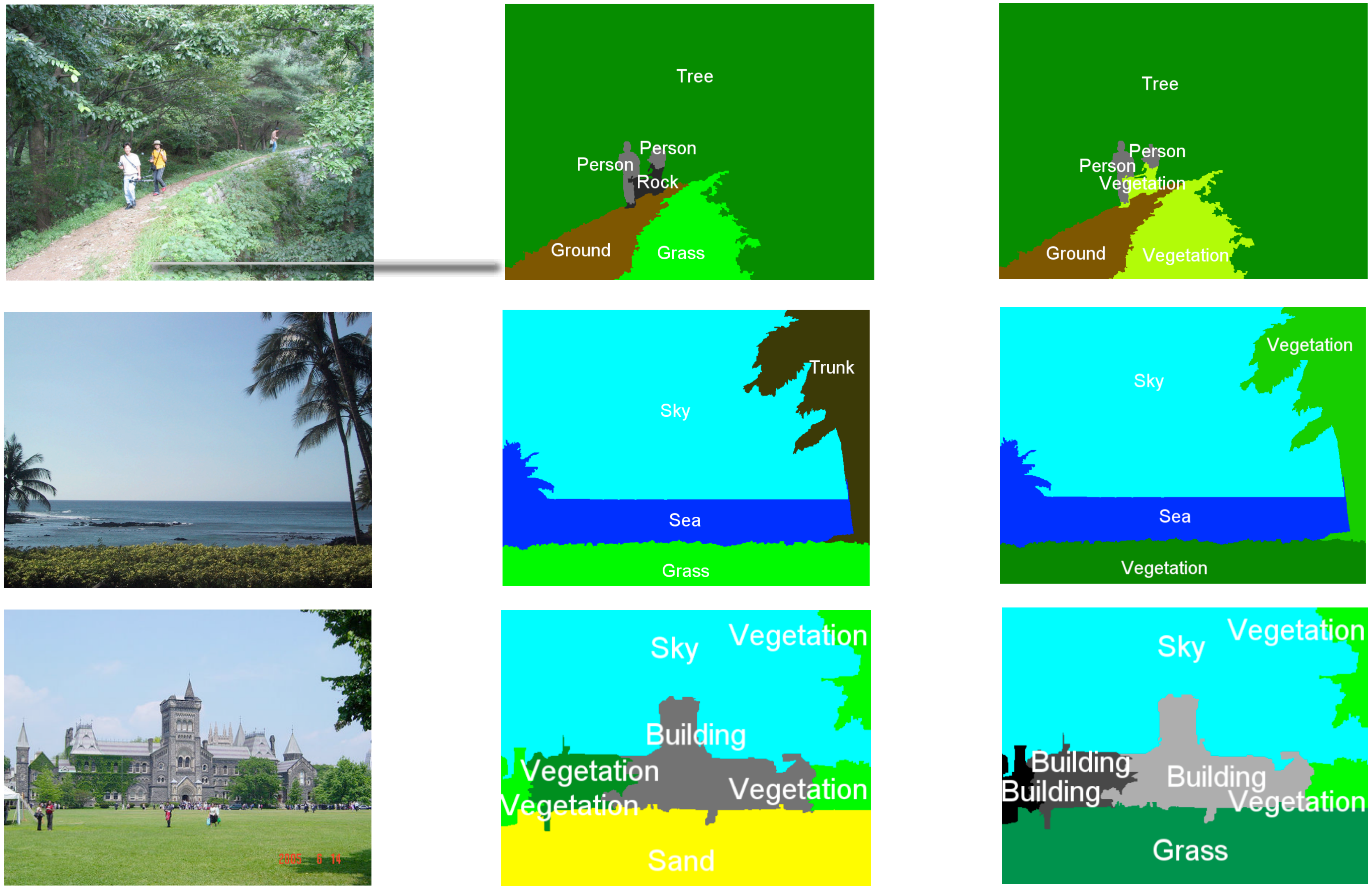
Concept Detection



Indicative image - sub-domain association results

Input Image		
Global Image Classification	Buildings: 0.12 Rockyside: 0.23 Forest: 0.42 Seaside: **0.84**	Buildings: 0.14 Rockyside: **0.52** Forest: 0.33 Seaside: 0.09
Region-based Image Classification	Buildings: 0.14 Rockyside: 0.22 Forest: 0.31 Seaside: **0.79**	Buildings: 0.23 Rockyside: 0.36 Forest: **0.38** Seaside: 0.14
Final Image Classification	**Seaside**	**Rockyside**

Indicative region-concept association results



Input image

Initial region-concept association

Final region-concept association

Formal Reasoning Approaches

- Logic-based approaches
  - Extensions of formal theories
  - Ad-hoc solutions based on crisp reasoners
- Statistical approaches
  - Bayesian Networks
- Support of imprecision - uncertainty

Reasoning

- Used for
  - Fusion
  - Consistency checking
  - Higher-level results

Reasoning

- Annotations from different analysis modules, of varying granularities (scene/region level)
  - partially contradictory
  - partially overlapping
  - restricted abstraction
- Need to be integrated into a semantically coherent overall representation

Proposed Solution

- OWL DL Ontologies
  - utilize domain semantics to integrate/enhance/remove annotations
- Fuzzy extensions on top of crisp DL reasoning to benefit from analysis extracted confidence values
- (DL-safe) Rules to address role value map expressivity requirements

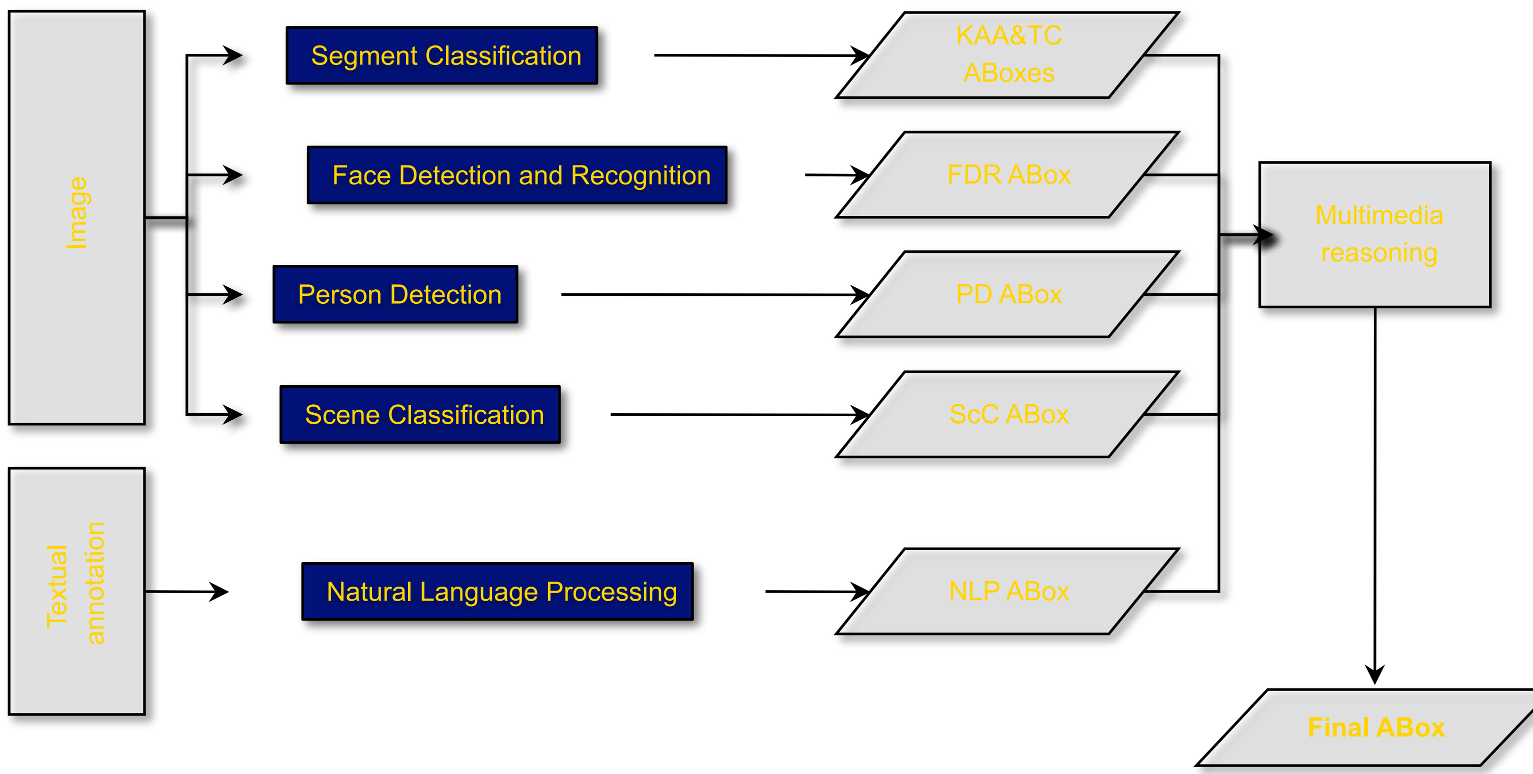
Reasoning Tasks

- Domain selection
  - input: direct and inferred (from region based ones) scene level instances
  - confidence values propagation along scene concept hierarchy
  - average of the (provenance-based) weighted annotations extracted per module
- Consistency checking
  - scene to scene annotations
  - scene to region annotations
  - removal of instances that triggered inconsistent inferences
- Make implicit inferences explicit
  - annotation enrichment

Architecture

$$\mathit{Natural} \equiv \mathit{Outdoors} \sqcup \neg\, \mathit{ManMade}$$
$$\mathit{Cityscape} \sqsubseteq \mathit{Manmade}$$
$$\mathit{Beach} \equiv \exists \text{contains}.\mathit{Sea} \sqcap \exists \text{contains}.\mathit{Sand}$$
$$\text{Cityscape} \equiv \ \$\ \text{contains.Building} \sqcap \$\text{contains}.\mathit{Road}$$
$$\mathit{Beach} \sqsubseteq \mathit{Natural}$$

TBox



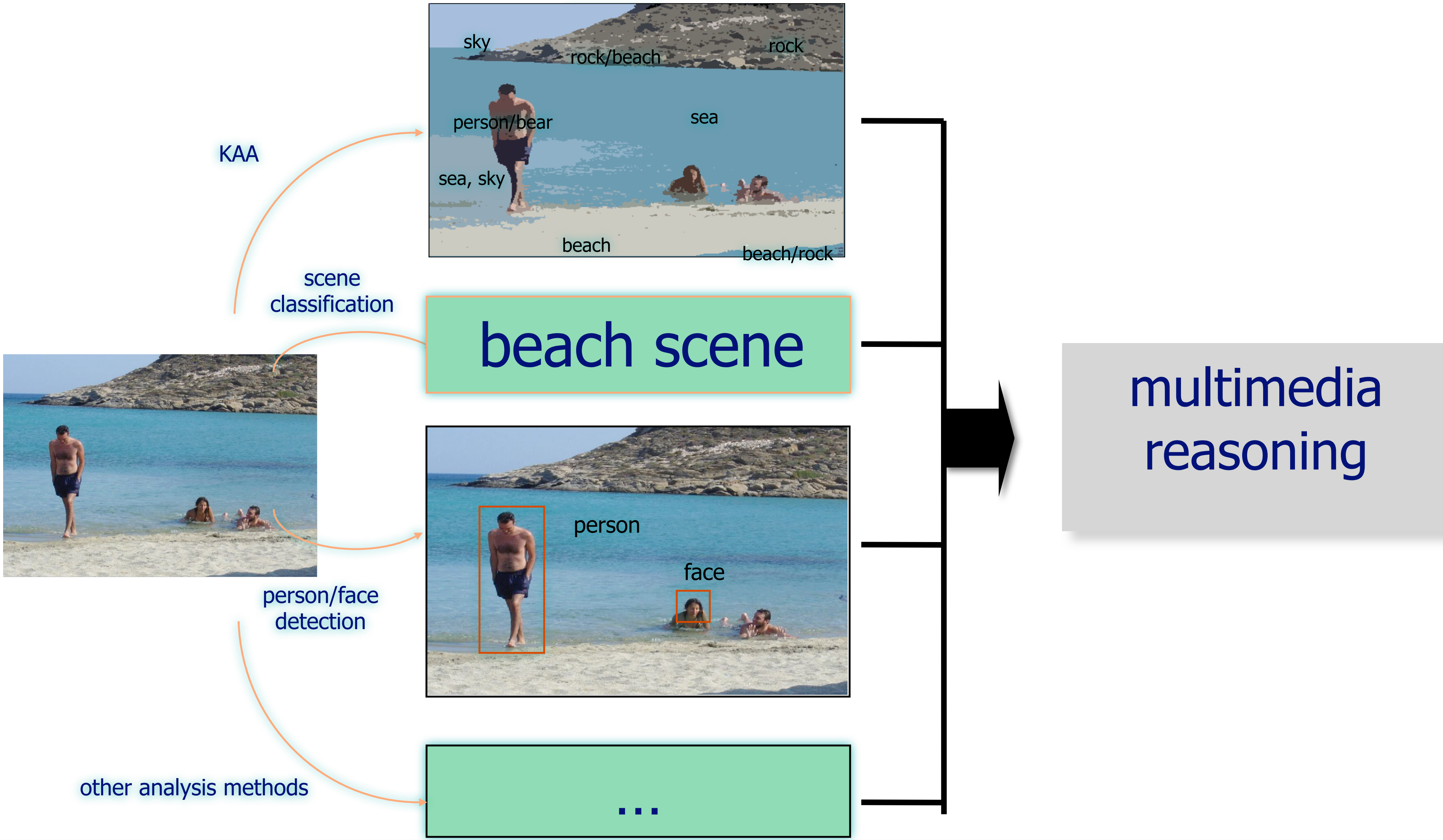
ABoxes:

analysis modules

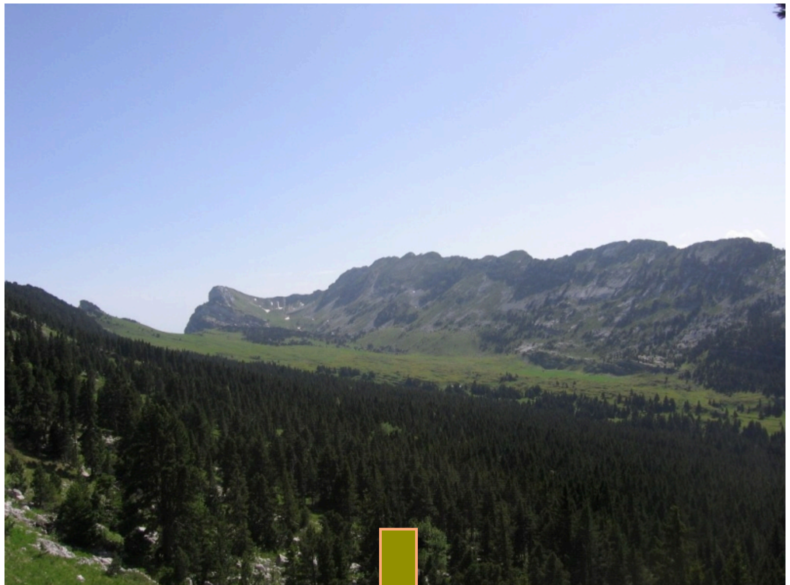
annotations

graded labels

Context and Reasoning for Analysis, aceMedia



Successive Reasoning



Final ABox'

Mountain >= 1.0 Green>= 1.0
Conifers >= 0.8

Beach, Sea, Sand, Cliff, Building: removed

2nd MMR

NLP ABox

Mountain >= 1.0
Green >= 1.0

(textual annotation added)

KAA ABox

Sand >= 0.79
Sky >= 0.75
Sea >= 0.75
Person >= 0.71

TC ABox

Building >= 0.27
Beach >= 0.53
Foliage >= 0.50
Cliff >= 0.62
Conifers >= 0.80

VCD ABox

Beach – 0.89

1st MMR

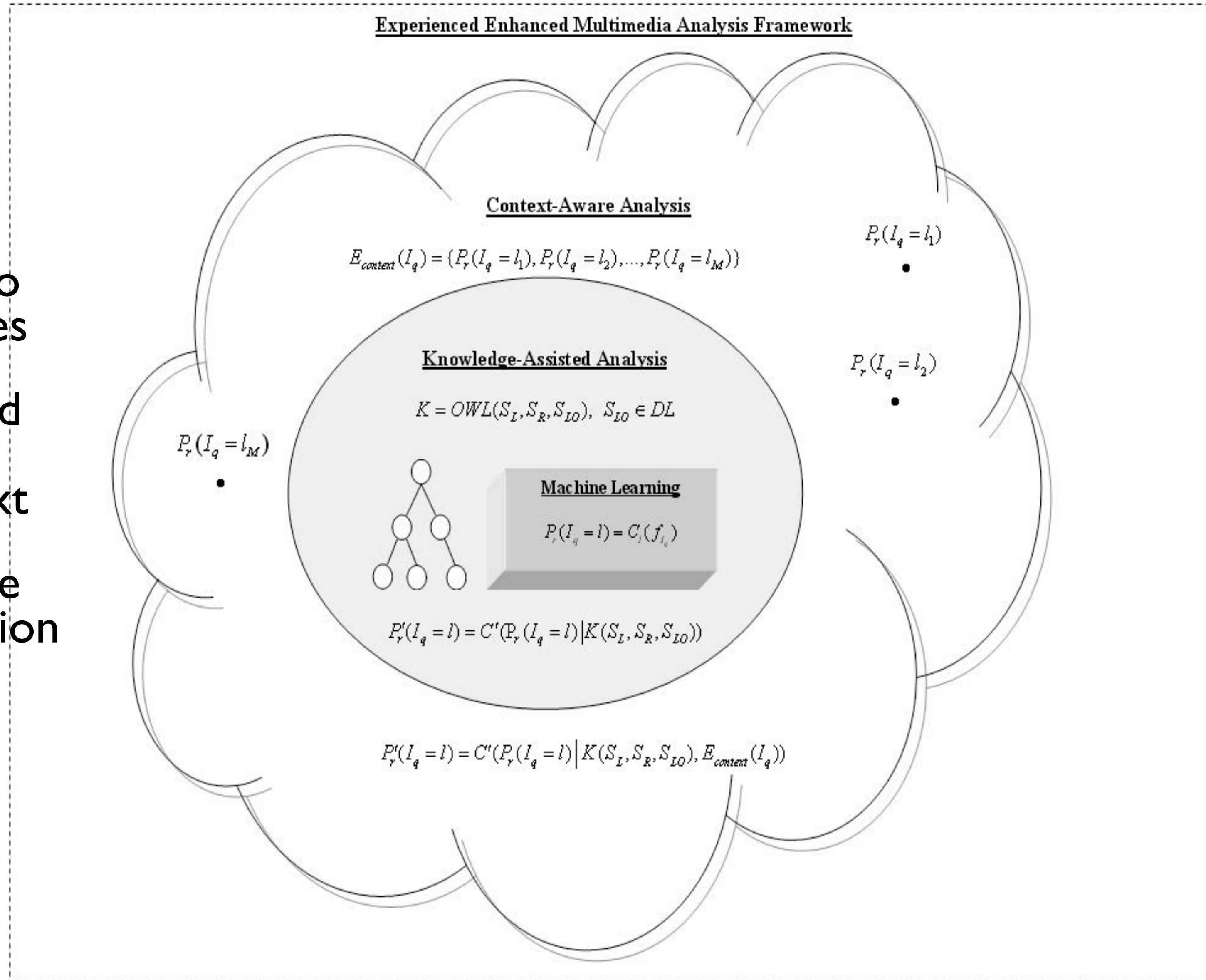
Final ABox

Beach >= 0.72 Sand >= 0.79
Sea >= 0.71 Sky
Foliage >= 0.50 Clif

Building, Conifers: removed

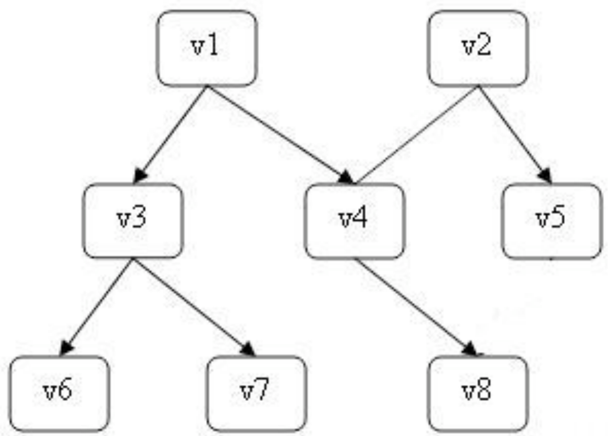
Reasoning with Bayesian Nets

- Machine learning to extract probabilities
- Knowledge-assisted analysis to define domain and context
- Perform knowledge inference for decision making using a Bayesian Network

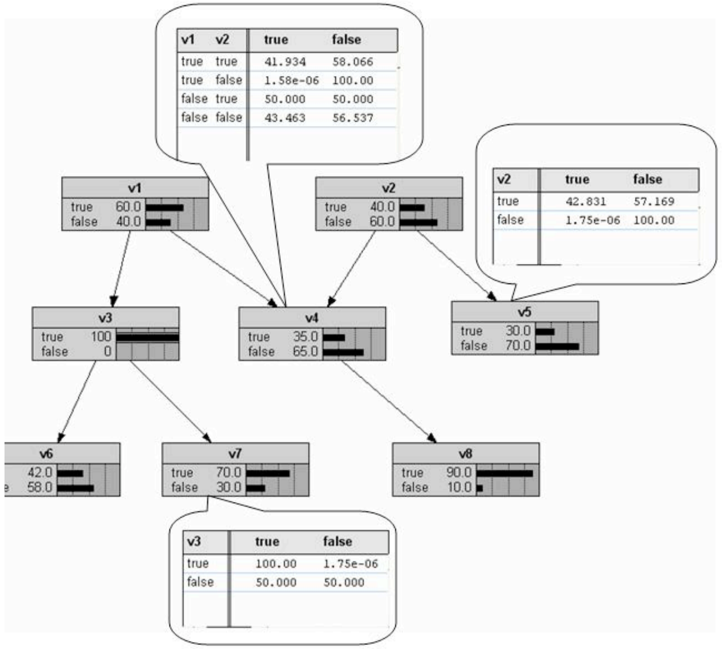


Bayesian Network Modeling

- Network Structure
  - Number of nodes
  - Placement of arcs



Parameter learning
Prior probabilities
Conditional Probability
Tables



Ontology to BN mapping

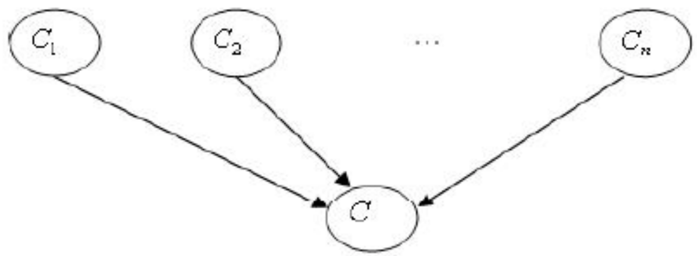


Fig.1. – "rdfs:subClassOf"

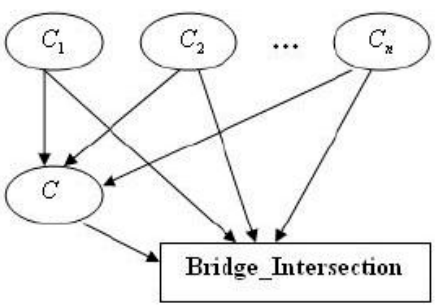


Fig.2. – "owl:intersectionOf"

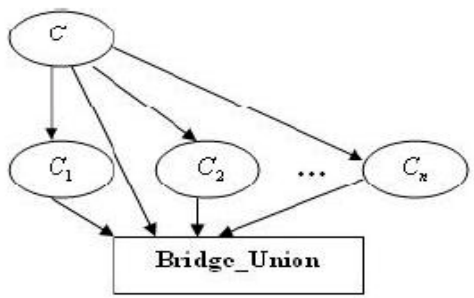
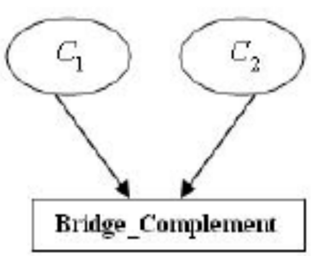
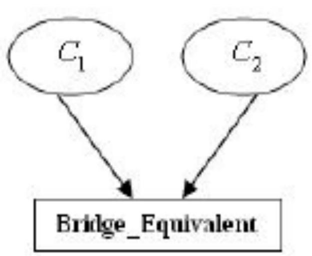
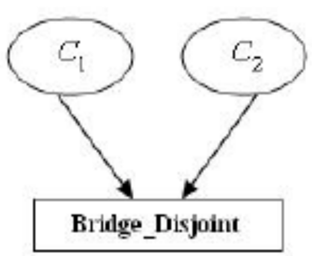


Fig.3. – "owl:unionOf"







(1) Bridge_Complement (Table 1): When its state is set to "True", C_1 and C_2 are complement of each other;

Table 1 – CPT of Bridge_Complement

C1	C2	True	False
True	True	0.000	100.00
True	False	100.00	0.000
False	True	100.00	0.000
False	False	0.000	100.00

(5) Bridge_Union (Table 5): When its state is set to "True", C is the union of C_1 and C_2;

Table 5 – Bridge_Union

C1	C2	C	True	False
True	True	True	100.00	0.000
True	True	False	0.000	100.00
True	False	True	100.00	0.000
True	False	False	0.000	100.00
False	True	True	100.00	0.000
False	True	False	0.000	100.00
False	False	True	0.000	100.00
False	False	False	100.00	0.000

(2) Bridge_Disjoint (Table 2): When its state is set to "True", C_1 and C_2 are disjoint with each other;

Table 2 – CPT of Bridge_Disjoint

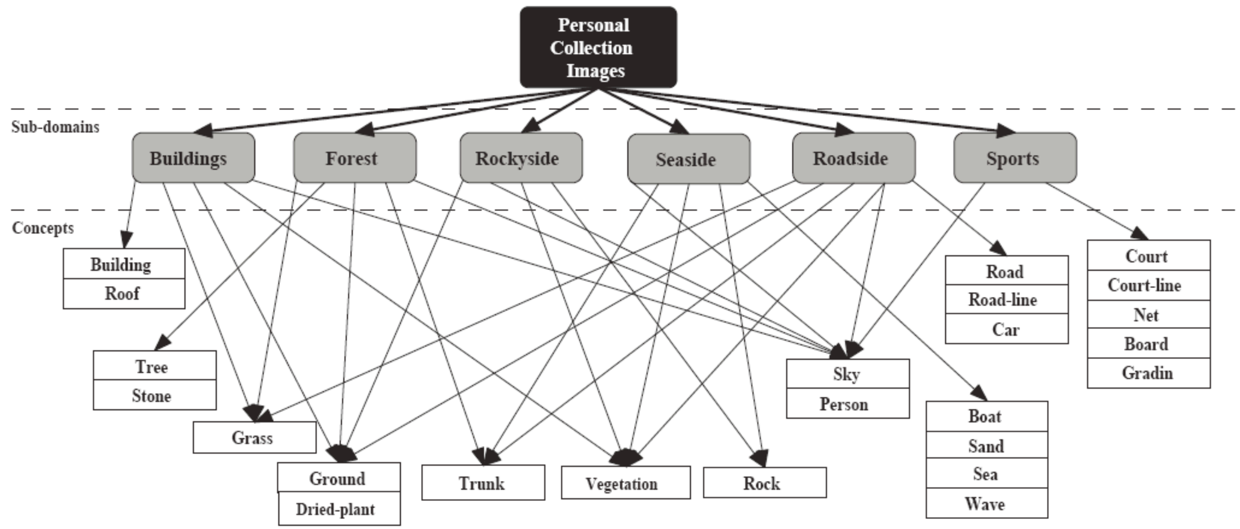
C1	C2	True	False
True	True	0.000	100.00
True	False	100.00	0.000
False	True	100.00	0.000
False	False	100.00	0.000

(4) Bridge_Intersection (Table 4): When its state is set to "True", C is the intersection of C_1 and C_2;

Table 4 – Bridge_Intersection

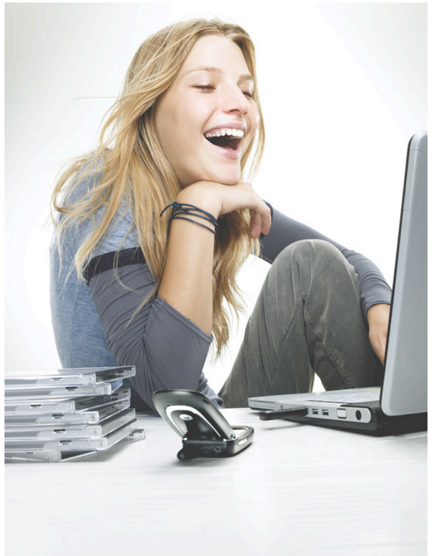
C1	C2	C	True	False
True	True	True	100.00	0.000
True	True	False	0.000	100.00
True	False	True	0.000	100.00
True	False	False	100.00	0.000
False	True	True	0.000	100.00
False	True	False	100.00	0.000
False	False	True	0.000	100.00
False	False	False	100.00	0.000

Preliminary Results



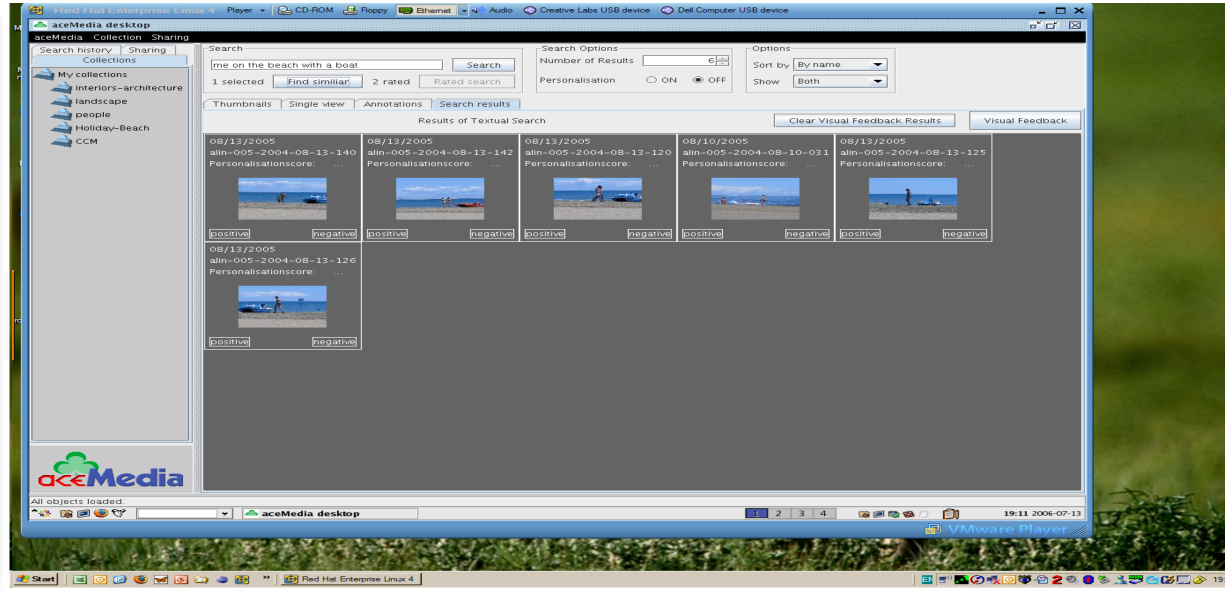
		Tennis	Roadside	Rockyside	Seaside	Forest	Countryside_Buildings	Average
Global classifiers solely	accuracy	96,91	**89,19**	90,43	90,43	87,65	89,19	90,63
	recall	**100,00**	68,42	**68,62**	85,71	76,66	30,00	71,57
	precision	83,33	**69,64**	70,00	67,60	63,88	100,00	75,74
	f-measure	90,90	**69,02**	69,30	75,59	69,69	46,15	70,11
Global classifiers with backgound knowledge	accuracy	**98,76**	88,27	87,96	90,43	83,02	88,58	89,50
	recall	100,00	66,66	41,17	85,71	**86,66**	26,00	67,70
	precision	**92,59**	66,66	70,00	67,60	52,52	**100,00**	74,90
	f-measure	**96,15**	66,66	51,85	75,59	65,40	41,26	66,15
Classification using regional information only	accuracy	95,06	83,95	86,41	84,56	87,65	90,12	87,96
	recall	94,00	68,42	47,05	69,64	55,00	50,00	64,02
	precision	78,33	53,42	58,53	54,16	71,13	78,13	65,62
	f-measure	85,45	60,00	52,17	60,93	62,26	60,97	63,63
Global classifiers with backgound knowledge and regional information	accuracy	98,14	88,27	**91,35**	**91,97**	**89,50**	**91,97**	**91,87**
	recall	98,00	**73,68**	64,70	**91,07**	71,66	**54,00**	**75,52**
	precision	90,74	64,61	**76,74**	**70,83**	**71,66**	90,00	**77,43**
	f-measure	94,23	68,85	**70,21**	**79,68**	**71,66**	**67,50**	**75,36**

aceMedia applications

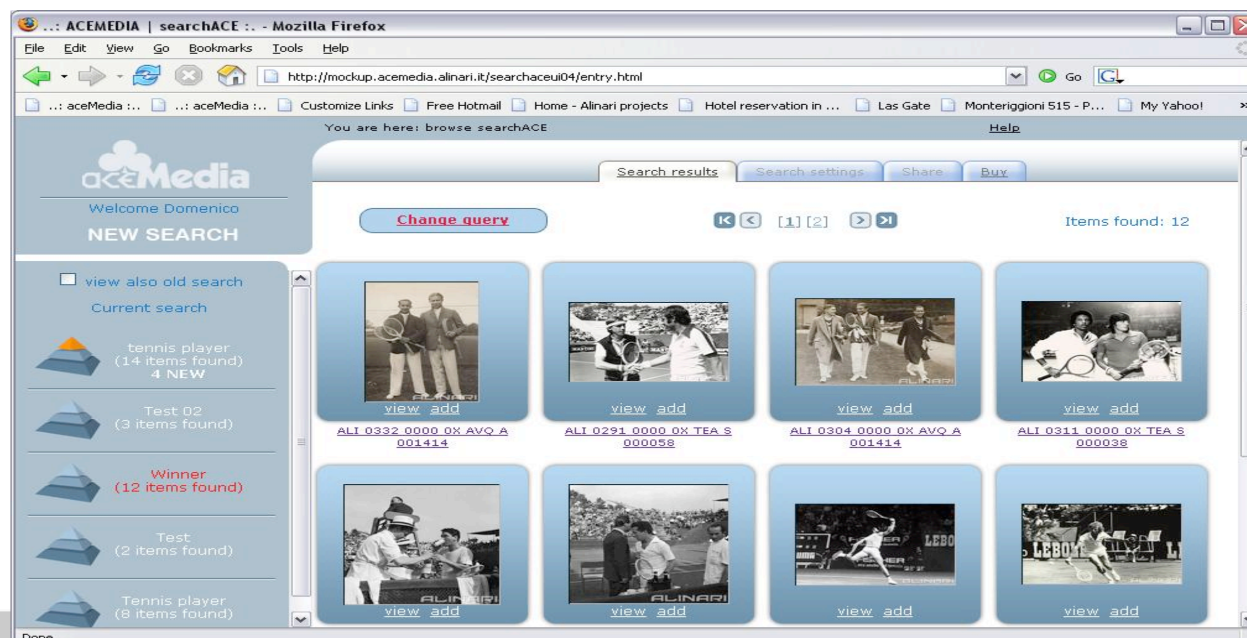


aceMedia PC applications

Standalone



Web-based



Common (Open) Issues

- Evaluation
- Annotated content
- Ontologies
- Fusion in analysis
- Uncertainty in reasoning
- Large-Scale
- Generic vs. Specific approaches
- Multiple domains support

Dissemination Activities

- SMART: Semantic MultimediA Research and Technology, networking cluster
- SAMT: International Conference on Semantics and digital Media Technologies (EWIMT)
  - 2007: 5-7 December 2007, Genova, Italy
- SSMS: Summer School on Multimedia Semantics
  - 2008: Crete, September

Conclusions

- Semantic analysis of multimedia is already providing results
- Fundamental and applied research in
  - Logic-based + signal approaches
  - Implicit + explicit (knowledge) approaches
- Different applications and requirements
- Ongoing research in all areas
- Future direction: analysis+reasoning for social (Web 2.0) applications

Thank You!

Multimedia Knowledge Lab, http://mklab.iti.gr