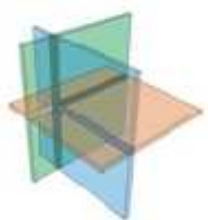


# Semantic Web

## - Multimedia Annotation -

Steffen Staab

<http://isweb.uni-koblenz.de>



UNIVERSITÄT  
KOBLENZ · LANDAU



## Different levels of annotations

- ◆ Metadata
  - Often technical metadata
  - EXIF, Dublin Core, access rights
- ◆ Content level
  - Semantic annotations
  - Keywords, domain ontologies, free-text
- ◆ Multimedia level
  - low-level annotations
  - Visual descriptors, such as dominant color

refers to information about technical details

creation details

- ◆ creator, creationDate, ...
- ◆ Dublin Core

camera details

- ◆ settings
- ◆ resolution
- ◆ format
- ◆ EXIF

access rights

- ◆ administrated by the OS
- ◆ owner, access rights, ...

Describes what is depicted and directly perceivable by a human  
usually provided manually

- ◆ keywords/tags
- ◆ classification of content

seldom generated automatically

- ◆ scene classification
- ◆ object detection

different types of annotations

- ◆ global vs. local
- ◆ different semantic levels

Global annotations most widely used

- ◆ flickr: tagging is only global
- ◆ organization within categories
- ◆ free-text annotations
- ◆ provide information about the content as a whole
- ◆ no detailed information

Local annotations are less supported

- ◆ e.g. flickr, PhotoStuff allow to provide annotations of regions
- ◆ especially important for semantic image understanding
  - allow to extract relations
  - provide a more complete view of the scene
- ◆ provide information about different regions
- ◆ and about the depicted relations and arrangements of objects

Free-Text annotations cover large aspects, but less appropriate for sharing, organization and retrieval

- ◆ Free-Text Annotations probably most natural for the human, but provide least formal semantics

Tagging provides light-weight semantics

- ◆ Only useful if a fixed vocabulary is used
- ◆ Allows some simple inference of related concepts by tag analysis (clustering)
- ◆ No formal semantics, but provides benefits due to fixed vocabulary
- ◆ Requires more effort from the user

Ontologies

- ◆ Provide syntax and semantic to define complex domain vocabularies
- ◆ Allow for the inference of additional knowledge
- ◆ Leverage interoperability
- ◆ Powerful way of semantic annotation, but hardly comprehensible by “normal users”

## Web-based Tools

- ◆ flickr
- ◆ riya

## Stand-Alone Tools

- ◆ PhotoStuff
- ◆ AktiveMedia

## Annotation for Feature Extraction

- ◆ M-OntoMat-Annotizer

Web2.0 application

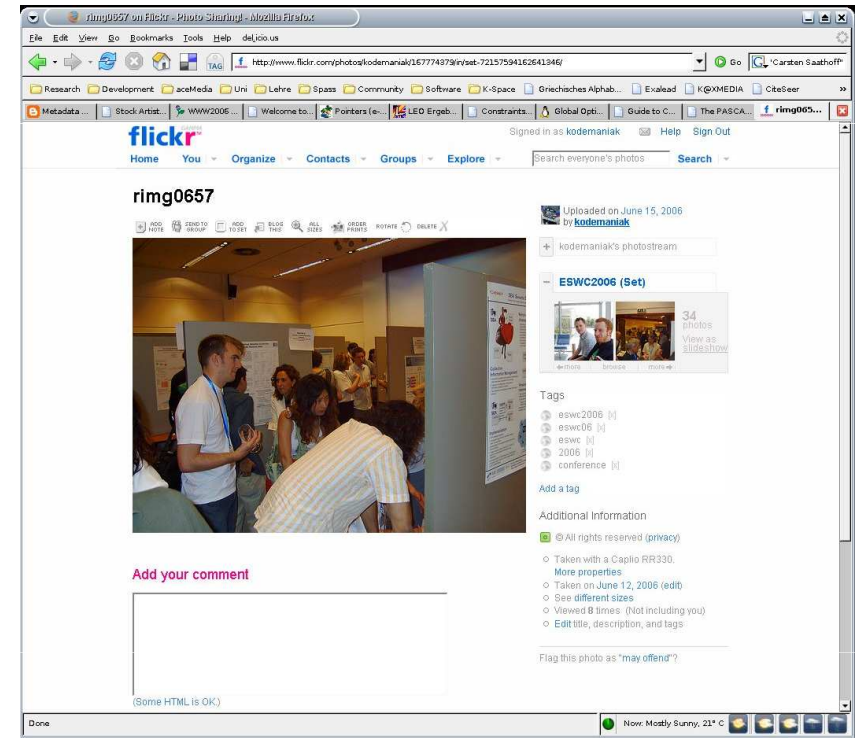
tagging photos globally

add comments to image regions  
marked by bounding box

large user community and tagging  
allows for easy sharing of images

partly fixed vocabularies evolved

- ◆ e.g. Geo-Tagging

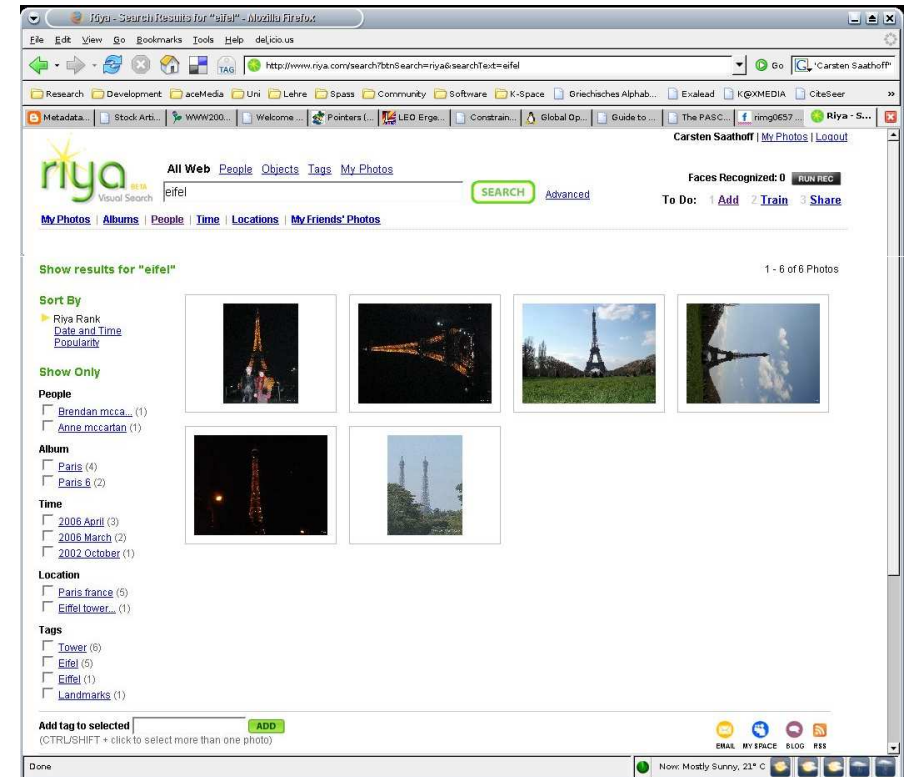


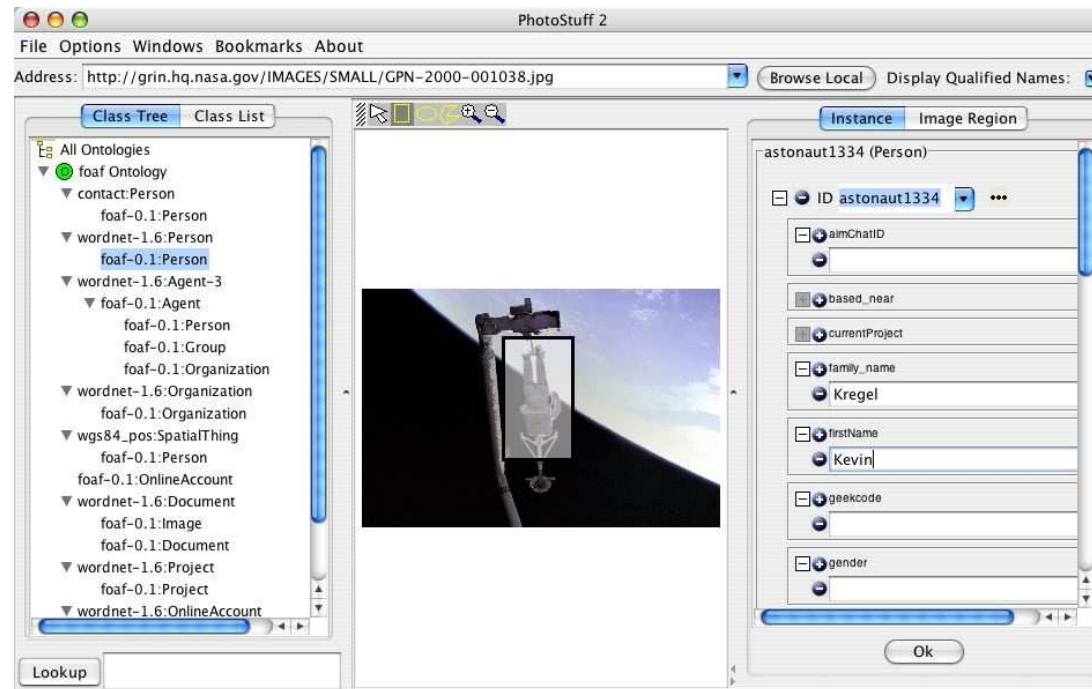
Similar to flickr in functionality

Adds automatic annotation features

- ◆ Face Recognition

- Mark faces in photos
- associate name
- train system
- automatic recognition of the person in the future





Java application for the annotation of images and image regions with domain ontologies

Used during ESWC2006 for annotating images and sharing metadata

Developed within Mindswap

Text and image annotation tool

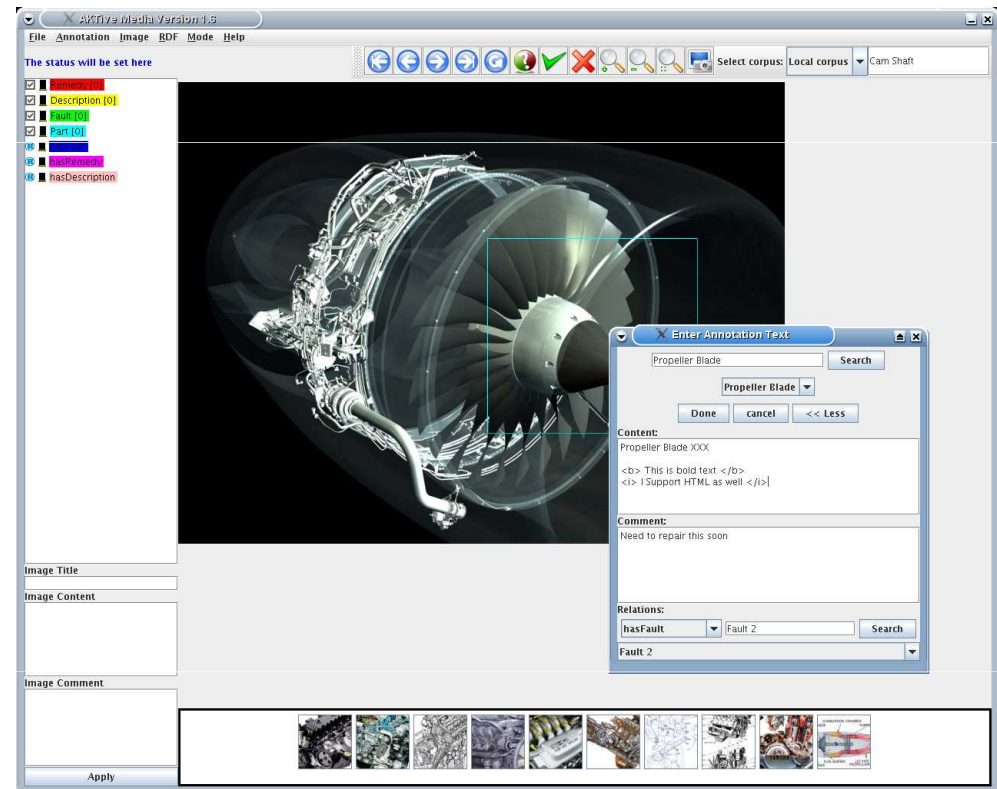
Region-based annotation

Uses ontologies

- ◆ suggests concepts during annotation
- ◆ providing a simpler interface for the user

Provides semi-automatic annotation of content, using

- ◆ Context
- ◆ Simple image understanding techniques
- ◆ flickr tagging data



Extracts knowledge from image regions for automatic annotation of images

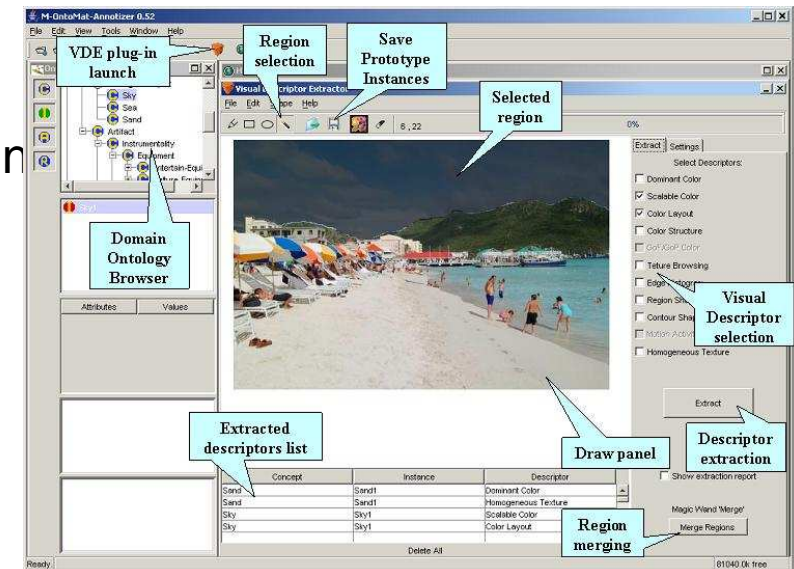
Extracting features:

- ◆ User can mark image regions manually or using an automatic segmentation tool
- ◆ MPEG-7 descriptors are extracted
- ◆ Stored within domain ontologies as prototypical, visual knowledge

Developed within aceMedia

Currently Version 2 is under development, incorporating

- ◆ true image annotation
- ◆ central storage
- ◆ extended knowledge extraction
- ◆ extensible architecture using a high-level multimedia ontology



Semantic annotation of images requires multimedia ontologies

- ◆ several vocabularies exist (Dublin Core, FOAF)
- ◆ they don't provide appropriate models to describe multimedia content sufficiently for sophisticated applications

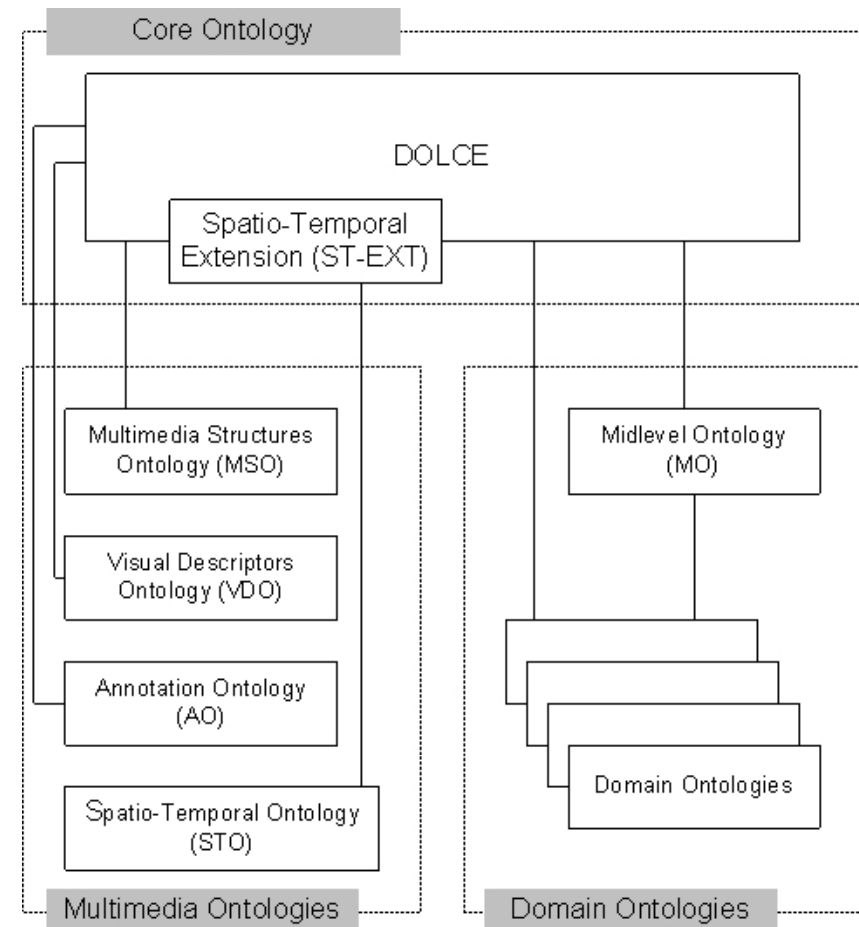
MPEG-7 provides an extensive standard, but especially semantic annotations are insufficiently supported

Several mappings of MPEG-7 into RDF or OWL exist

- ◆ now: VDO and MSO developed within aceMedia
- ◆ later: Engineering a multimedia upper ontology

## aceMedia Multimedia Ontology Infrastructure

- ◆ DOLCE as core ontology
- ◆ Multimedia Ontologies
  - Visual Descriptors Ontology (VDO)
  - Multimedia Structures Ontology (MSO)
  - Annotation and Spatio-Temporal Ontology augmenting VDO and MSO
- ◆ Domain Ontologies
  - capture domain specific knowledge



## Representation of MPEG-7 Visual Descriptors in RDF

- ◆ Visual Descriptors represent low-level features of multimedia content
- ◆ e.g. dominant color, shape or texture

## Mapping to RDF allows for

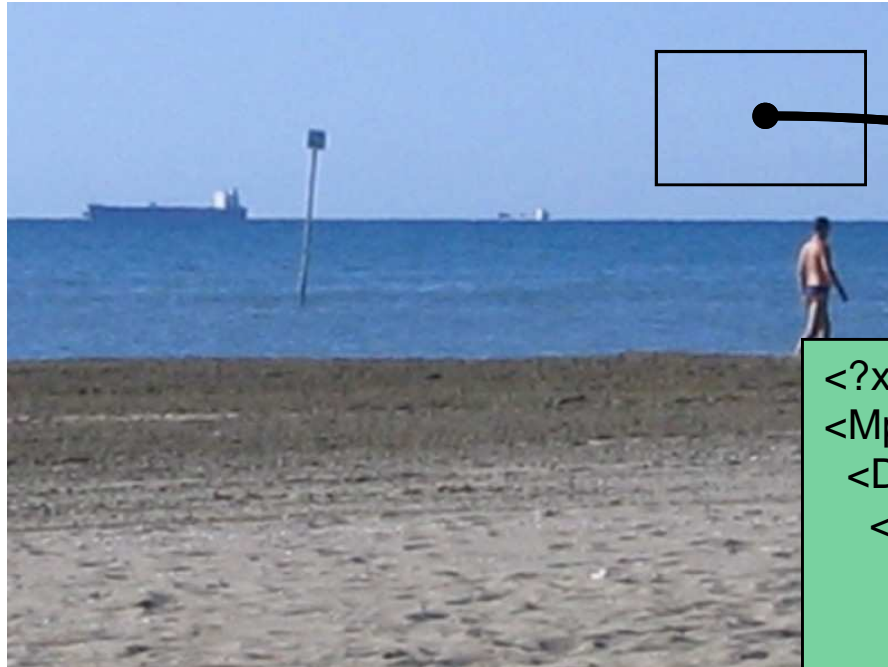
- ◆ linking of domain ontology concepts with visual features
- ◆ better integration with semantic annotations
- ◆ a common underlying model for visual and semantic features

Used for automatic annotation of images

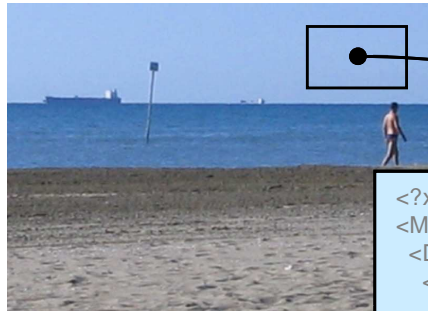
Idea:

- ◆ Describe the visual appearance of domain concepts by providing examples
- ◆ User annotates instances of concepts and extracts features
- ◆ features are represented with the VDO
- ◆ the examples are then stored in the domain ontology as prototype instances of the domain concepts

Thus the names: prototype and prototypical knowledge



```
<?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>
```



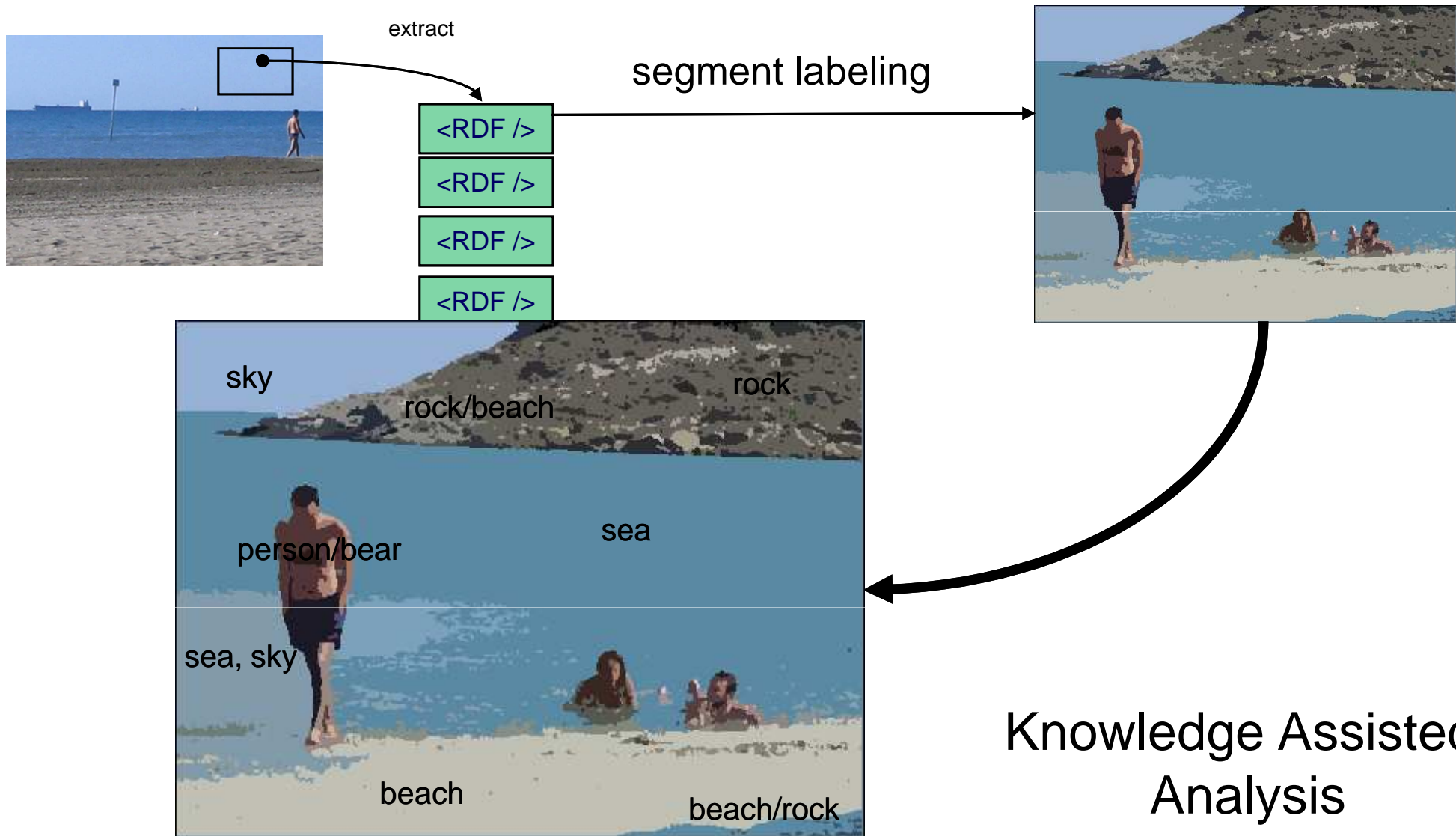
extract

```
<?xml version='1.0' encoding='ISO-8859-1'>
<Mpeg7 xmlns:scd="http://www.d3m.de/mpeg7/scd" xmlns:ext="http://www.d3m.de/mpeg7/ext">
  <DescriptionUnit xsi:type = "DescriptionUnit" >
    <Descriptor xsi:type = "DominantColorDescriptor" >
      <SpatialCoherency>31</SpatialCoherency>
      <Value>
        <Percentage>31</Percentage>
        <Index>19 23 29</Index>
        <ColorVariance>0 0 0</ColorVariance>
      </Value>
    </Descriptor>
  </DescriptionUnit>
</Mpeg7>
```

transform

```
<vdo:ScalableColorDescriptor rdf:ID="vde-inst1">
  <vdo:coefficients> 0 [...] 1 </vdo:coefficients>
  <vdo:numberOfBitPlanesDiscarded> 6
</vdo:numberOfBitPlanesDiscarded>
  <vdo:numberOfCoefficients> 0
</vdo:numberOfCoefficients>
</vdo:ScalableColorDescriptor>

<vdoext:Prototype rdf:ID="Sky Prototype 1">
  <rdf:type rdf:resource="#Sky"/>
  <vdoext:hasDescriptor
    rdf:resource="#vde-inst1"/>
</vdoext:Prototype>
```



RDF representation of the MPEG-7 Multimedia Description Schemes

Contains only classes and relations relevant for representing a decomposition of images or videos

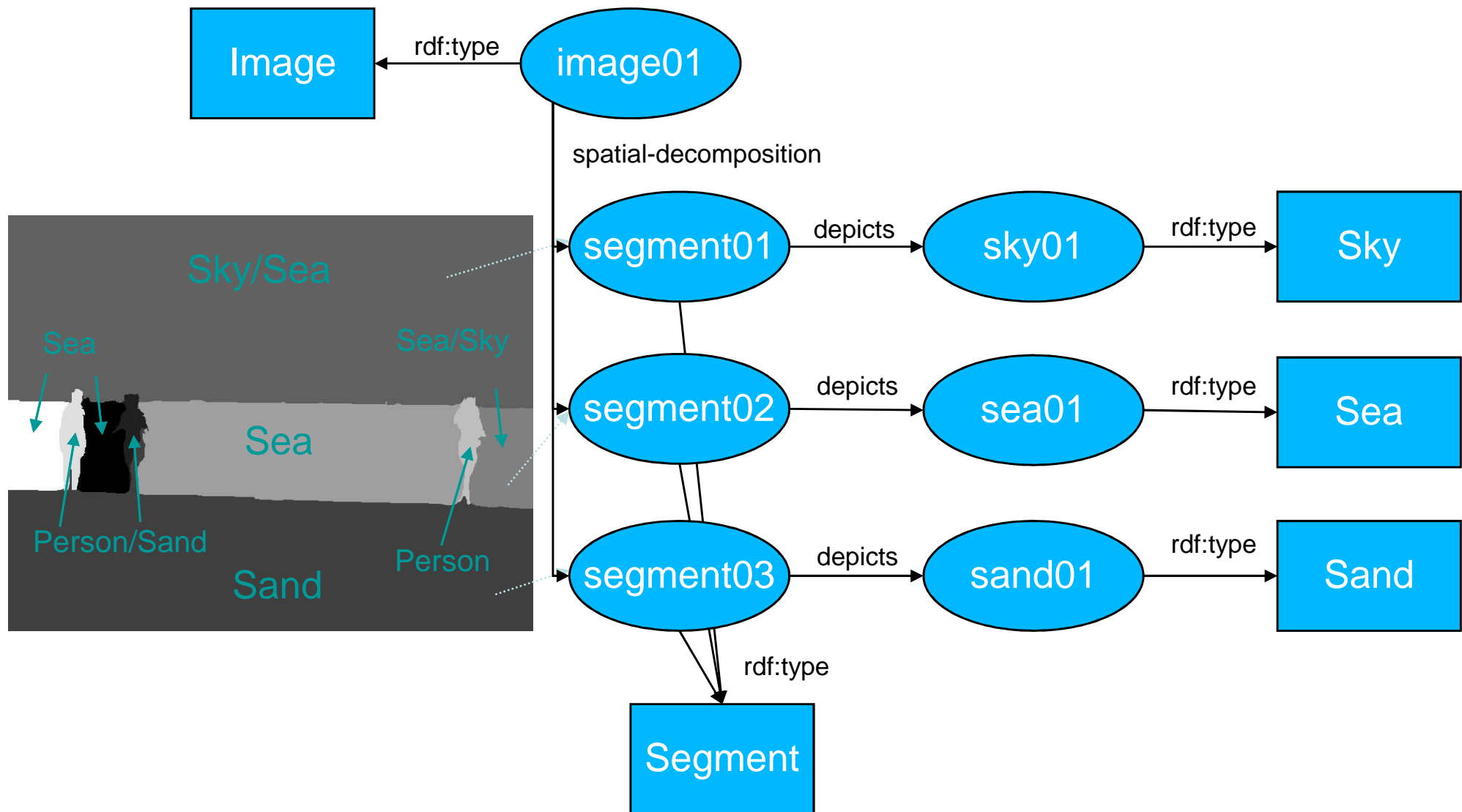
Contains Classes for different types of segments

- ◆ temporal and spatial segments

Contains relations to describe different decompositions

Augmented by annotation ontology and spatio-temporal ontology, allowing to describe

- ◆ regions of an image or video
- ◆ the spatial and temporal arrangement of the regions
- ◆ what is depicted in a region



# Thank you

Acknowledgements to Carsten Saathoff @ ISWeb for majority of slides