Knowledge-Driven Video Information Retrieval with LOD

Outline

- Video Retrieval Challenges and Limitations
- Unstructured vs. Structured Annotations
- Bridging the Semantic Gap
- Multimedia Vocabularies and Ontologies
- Linked (Open) Data for multimedia
- Standardization of Video Annotations
Bridging the Semantic Gap
Unstructured vs. Structured Data

Plain text tags
XML/XSD metadata
RDFS/OWL annotations
Bridging the Semantic Gap

Vocabularies and Ontologies

- Dublin Core
- Creative Commons
- Schema.org

Concept definitions
Classes, properties, individuals
Relationships
Rules
Video Object Representation
Lightweight Annotations

- HTML5 Microdata
- JSON-LD
- RDFa
- Microformats
Video Object Representation
Lightweight Annotations

<video><source src="friday.mp4" type="video/mp4" /></video>
Directed by <span itemprop="director">John Smith</span>
Video Object Representation
Lightweight Annotations

<script type="application/ld+json">
{
  "@context": "http://schema.org",
  "video": {
    "@type": "VideoObject",
    "name": "The Koala Hug",
    "duration": "T1M48S"
  }
}
</script>
## Video Object Representation

### Semi-Structured Vocabularies

<table>
<thead>
<tr>
<th>Vocabulary</th>
<th>Standard</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPEG-7</td>
<td>ISO/IEC 15938</td>
<td>XSD</td>
</tr>
<tr>
<td>MPEG-21</td>
<td>ISO/IEC 21000</td>
<td>XSD</td>
</tr>
<tr>
<td>NewsML</td>
<td>IPTC NewsML-G2</td>
<td>XSD</td>
</tr>
<tr>
<td>TV-Anytime</td>
<td>ETSI TS 102 822</td>
<td>XSD</td>
</tr>
</tbody>
</table>
### Video Object Representation

Semi-Structured Vocabularies

<table>
<thead>
<tr>
<th>Vocabulary</th>
<th>Standard</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPEG-7</td>
<td>ISO/IEC 15938</td>
<td>XSD</td>
</tr>
<tr>
<td>MPEG-21</td>
<td>ISO/IEC 21000</td>
<td>XSD</td>
</tr>
<tr>
<td>NewsML</td>
<td>IPTC NewsML-G2</td>
<td>XSD</td>
</tr>
<tr>
<td>TV-Anytime</td>
<td>ETSI TS 102 822</td>
<td>XSD</td>
</tr>
</tbody>
</table>
Bridging the Semantic Gap
XSD to RDFS/OWL Mapping

MPEG-7

COMM

MPEG-7Ontos

SWIntO

…
Bridging the Semantic Gap
Issues Inherited from MPEG-7

• Strong focus on low-level descriptors

“color distribution feature values of an image for red, black, and yellow still do not allow the conclusion that the image shows a sunset”

Boll et al., 1998
Knowledge-Driven Video Information Retrieval with LOD
Bridging the Semantic Gap
Issues Inherited from MPEG-7

Knowledge-Driven Video Information Retrieval with LOD

ESAIR'15
Bridging the Semantic Gap
Issues Inherited from MPEG-7

<owl:Class rdf:about="#cbac-crac-coefficient-27-descriptor-parameter"/>
<owl:Class rdf:about="#nested-visual-descriptor-parameter"/>
Bridging the Semantic Gap
Issues Inherited from MPEG-7

<owl:Class rdf:about="#cbac-crac-coefficient-27-descriptor-parameter"/>
<rdfs:subClassOf>
<owl:Class rdf:about="#nested-visual-descriptor-parameter"/>

Bridging the Semantic Gap

Issues Inherited from MPEG-7

- Strong focus on low-level descriptors
- Conceptual ambiguity
- Semantic interoperability issues
- Syntactic interoperability issues
- Structural complexity: 1,182 elements, 417 attributes, and 377 complex types
Bridging the Semantic Gap
Custom Ontologies

- Large Scale Concept Ontology for Multimedia (LSCOM)
- Linked Movie Database Ontology
- Multimedia Metadata Ontology (M3O)
- Ontology for Media Resources
  ...

Knowledge-Driven Video Information Retrieval with LOD  ESAIR’15
Bridging the Semantic Gap
Custom Ontologies

• T TBox: terminological knowledge
• A ABox: assertional knowledge
Bridging the Semantic Gap

Custom Ontologies

- TBox: terminological knowledge
- ABox: assertional knowledge

Knowledge Base

Knowledge-Driven Video Information Retrieval with LOD
Bridging the Semantic Gap

Custom Ontologies

- $\mathcal{T}$ TBox: terminological knowledge
- $\mathcal{A}$ ABox: assertional knowledge
- $\mathcal{R}$ RBox: role inclusion axioms + transitivity axioms
## Multimedia Ontologies

Limited DL Expressivity

<table>
<thead>
<tr>
<th>Ontology</th>
<th>Language</th>
<th>DL</th>
</tr>
</thead>
<tbody>
<tr>
<td>LinkedMDB</td>
<td>RDFS</td>
<td>$\mathcal{AL}$</td>
</tr>
<tr>
<td>LSCOM</td>
<td>OWL</td>
<td>$\mathcal{AL}$</td>
</tr>
<tr>
<td>M3O</td>
<td>OWL</td>
<td>$\mathit{SHIQ}^{(D)}$</td>
</tr>
<tr>
<td>COMM</td>
<td>OWL</td>
<td>$\mathit{SHOIN}^{(D)}$</td>
</tr>
</tbody>
</table>
# Multimedia Ontologies

## Full DL Expressivity

<table>
<thead>
<tr>
<th>Ontology</th>
<th>Language</th>
<th>DL</th>
</tr>
</thead>
<tbody>
<tr>
<td>VidOnt</td>
<td>OWL 2</td>
<td>$SROIQ^D$</td>
</tr>
</tbody>
</table>

http://vidont.org
Bridging the Semantic Gap
Domain Ontologies

- OWL 2 ontologies + SWRL rules

Decidability
Bridging the Semantic Gap

Domain Ontologies

- OWL 2 ontologies + SWRL rules

Decidability

Solution: OWL 2 + DL Rules + DL-safe rules: very expressive & still decidable
Linked Data for Multimedia

Rationale

- Global identification with URIs
- Linking to annotations or media fragments in the Linked Data cloud
- Differentiate video objects and media fragments from information resources
- Support access through SPARQL queries
Video IR with LOD

Conclusions

• OWL 2 ontologies are needed
  a) Alignment with standards
  b) Exploit $SROIQ^D$ constructs
  c) Define Rbox axioms
  d) Use OWL 2+DL-safe rules

• LOD: video understanding, discovery, …

Advanced inference & reasoning support
Questions
Thank you.

For more on semantic multimedia, visit www.lesliesikos.com