Introduction to Description Logics and their usage in multimedia analysis

Stamatia Dasiopoulou
11 Nov 2009
Outline

- Description Logics (DLs)
  - Basics (syntax, semantics, inference services)
  - OWL as a DL language

- DLs and multimedia analysis
  - Content representation (briefly)
  - Knowledge-based extraction / interpretation of content semantics
What are Description Logics?

- A family of logic-based knowledge representation formalisms
  - Descendants of semantics networks, frame-based systems

- Distinguished by
  - Formal semantics
    - Decidable fragments of FOL
  - Inference
    - Sound & complete, highly optimized (implemented) algorithms
What does it mean to “be” a formal language?

- **Syntax**
  - What expressions form valid sentences

- **Semantics**
  - What is the meaning of the expressed sentences

- **Reasoning procedures**
  - How is implicit knowledge derived from the explicitly stated one
  
(e.g. If Socrates is human and every human is mortal, I can derive the fact Socrates is mortal)
DL Basics

- **Concept** names are equivalent to unary predicates
  - Interpreted as sets of objects e.g. Person, Student

- **Role** names are equivalent to binary predicates
  - Interpreted as binary relations on objects e.g. hasChild, likes

- **Individual** names equal constants e.g. Mary, John, India

- **Constructors**: concept and role forming operators
DL family

- A given DL is defined by the set of allowed constructors.

- Smallest propositionally closed DL is $\mathcal{ALC}$, concept constructed using:
  - $\cap$, $\cup$, $\neg$
  - restricted $\exists$ and $\forall$ quantifiers
  - atomic roles

E.g., Person all of whose children are either Doctors or have a child who is a Doctor:

$$\text{Person } \sqcap \forall \text{hasChild} . (\text{Doctor } \sqcap \exists \text{hasChild} . \text{Doctor})$$
DL family (cont.)

- **Additional letters** indicate other extension, e.g.:
  - H for role inclusion axioms (role hierarchy)
  - O for nominals (singleton classes)
  - I for inverse roles
  - N for number restrictions ($\geq nR$, $\leq nR$)
  - Q for qualified number restrictions ($\geq nR.C$, $\leq nR.C$)
KR architecture based on DLs

Knowledge Base

TBox (terminology)
- Man ⊑ Human ∩ Male
- Woman ⊑ Human ∩ -Male
- Parent ⊑ Human ∩ ∃ hasChild.Human

ABox (data/facts)
- John : Man
- (John, Mary) : has-child

Inference System

Applications
A DL **Knowledge base** $\mathcal{K}$ is a pair $\langle \mathcal{T}, \mathcal{A} \rangle$ where
- $\mathcal{T}$ is a set of “terminological” axioms (the TBox)
- $\mathcal{A}$ is a set of “assertional” axioms (the ABox)

**TBox axioms** are of the form:
$$C \sqsubseteq D, \ C \equiv D, \ R \sqsubseteq S, \ R \equiv S \text{ and } R^+ \sqsubseteq R$$
where $C$, $D$ concepts, $R$, $S$ roles, and $R^+$ set of transitive roles

**ABox axioms** are of the form:
$$x: C, \ (x, y): R$$
where $x, y$ are individual names, $C$ a concept and $R$ a role
DL semantics

- Semantics defined by interpretations

- An interpretation $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$, where
  - $\Delta^\mathcal{I}$ is the domain (a non-empty set)
  - $\cdot^\mathcal{I}$ is an interpretation function that maps:
    - Concept (class) name $A \rightarrow$ subset $A^\mathcal{I}$ of $\Delta^\mathcal{I}$
    - Role (property) name $R \rightarrow$ binary relation $R^\mathcal{I}$ over $\Delta^\mathcal{I}$
    - Individual name $i \rightarrow i^\mathcal{I}$ element of $\Delta^\mathcal{I}$
DL Semantics (cont.)

DL Semantics (cont.)

- Interpretation function \( \mathcal{I} \) extends to concept (and role) expressions

\[
\begin{align*}
(C \cap D) & \mathcal{I} = C^\mathcal{I} \cap D^\mathcal{I} \\
(C \cup D) & \mathcal{I} = C^\mathcal{I} \cup D^\mathcal{I} \\
(\neg C) & \mathcal{I} = \Delta^\mathcal{I} \setminus C^\mathcal{I} \\
\{x\} & \mathcal{I} = \{x^\mathcal{I}\} \\
(\exists R.C) & \mathcal{I} = \{x \mid \exists y.\langle x, y \rangle \in R^\mathcal{I} \land y \in C^\mathcal{I}\} \\
(\forall R.C) & \mathcal{I} = \{x \mid \forall y.\langle x, y \rangle \in R^\mathcal{I} \Rightarrow y \in C^\mathcal{I}\} \\
(\leq nR) & \mathcal{I} = \{x \mid \#\{y \mid \langle x, y \rangle \in R^\mathcal{I}\} \leq n\} \\
(\geq nR) & \mathcal{I} = \{x \mid \#\{y \mid \langle x, y \rangle \in R^\mathcal{I}\} \geq n\} \\
(R^-) & \mathcal{I} = \{(x, y) \mid (y, x) \in R^\mathcal{I}\}
\end{align*}
\]
DL knowledge base Semantics

- An interpretation $\mathcal{I}$ satisfies (is a model) a TBox axiom $A (\mathcal{I}^2 A)$:
  
  $\mathcal{I}^2 C \sqsubseteq D$ iff $C^\mathcal{I} \subseteq D^\mathcal{I}$
  
  $\mathcal{I}^2 C \equiv D$ iff $C^\mathcal{I} = D^\mathcal{I}$
  
  $\mathcal{I}^2 R \sqsubseteq S$ iff $R^\mathcal{I} \subseteq S^\mathcal{I}$
  
  $\mathcal{I}^2 R \equiv S$ iff $R^\mathcal{I} = S^\mathcal{I}$
  
  $\mathcal{I}^2 R^+ \sqsubseteq R$ iff $(R^\mathcal{I})^+ \subseteq R^\mathcal{I}$

- $\mathcal{I}$ satisfies a TBox $\mathcal{T} (\mathcal{I}^2 \mathcal{T})$ iff $\mathcal{I}$ satisfies every axiom $A$ in $\mathcal{T}$

- An interpretation $\mathcal{I}$ satisfies (models) an ABox axiom $A (\mathcal{I}^2 A)$:

  $\mathcal{I}^2 x:\!D$ iff $x^\mathcal{I} \in D^\mathcal{I}$

  $\mathcal{I}^2 \langle x, y \rangle:\!R$ iff $(x^\mathcal{I}, y^\mathcal{I}) \in R^\mathcal{I}$

- $\mathcal{I}$ satisfies an ABox $\mathcal{A} (\mathcal{I}^2 \mathcal{A})$ iff $\mathcal{I}$ satisfies every axiom $A$ in $\mathcal{A}$

- $\mathcal{I}$ satisfies an KB $\mathcal{K} (\mathcal{I}^2 \mathcal{K})$ iff $\mathcal{I}$ satisfies both $\mathcal{T}$ and $\mathcal{A}$
CWA vs OWA

- **Closed World Assumption**: assumes that the available information is complete
  - If an assertion cannot be derived then its negation is deduced

- **Open World Assumption**: absence of information means lack of information
  - The assertion holds in some models, and doesn’t hold in others
Example

- $A_E = \{ \text{hasChild(Iokaste,Edipus)},$
  \text{hasChild(Iokaste,Polyneikes),}$
  \text{hasChild(Edipus,Polyneikis),}$
  \text{hasChild(Polyneikis,Thesandros),}$
  \text{Patricide(Edipus), } \neg \text{Patricide(Thesandros)} \}\.$

$A_E | = \{ \text{Iokaste: } \exists \text{hasChild.}(\neg \text{Patricide} \exists \text{hasChild.}\neg \text{Patricide}) \}$

???
Inference Services

- Every logical formalism provides its own reasoning services.

- Description Logics (DLs) provide reasoning services for TBoxes, ABoxes and Knowledge Bases (TBoxes and ABoxes together).
Inference Services for TBoxes

- **Satisfiability:** A concept $C$ is *satisfiable* w.r.t. a TBox $T$ if there exists a model $I$ of $T$ such that $C^I \neq \emptyset$. Then $I$ is called a *model* of $C$.

- **Subsumption:** A concept $C$ is *subsumed* by a concept $D$ w.r.t. $T$, written $C \sqsubseteq D$, if for every model $I$ of $T$, $C^I \subseteq D^I$.

- **Equivalence:** Two concepts $C$ and $D$ are *equivalent* w.r.t. $T$, written $C \equiv D$, if for every model $I$ of $T$, $C^I \equiv D^I$.

- **Disjointness:** Two concepts $C$ and $D$ are *disjoint*, w.r.t. $T$, written $C \neq D$, if for every model $I$ of $T$, $C^I \neq D^I$.
Some examples

- Is $\text{Man} \sqcap \text{Man}$ satisfiable w.r.t. an empty TBox? No
- Is $\text{Man} \sqcap \text{Woman}$ satisfiable w.r.t. an empty TBox? Yes

- $\text{Woman} \sqsubseteq \text{Person}$ w.r.t. empty TBox?
  No! one can create an interpretation where $\text{Woman}^I \sqsubseteq \text{Person}^I$

- $\text{Man} \sqsubseteq \text{Person}$ w.r.t. $T = \{ \text{Person} \equiv \text{Man} \sqcap \text{Woman} \}$?
  Yes! in all models $I$ of $T$, $\text{Person}^I$ contains all objects of $\text{Man}^I$ plus all of $\text{Woman}^I$. 
Inference Services for ABoxes

- **Consistency**: An ABox $A$ is *consistent* w.r.t. a TBox $T$ if there exists a model of $T$ which satisfies each assertion in $A$.

- **Entailment (Instance Checking)**: An ABox $A$ *entails* an assertion $\varphi$, written $A \models \varphi$ iff every interpretation that satisfies $A$ also satisfies the assertion.
Some examples

- Is $A = \{\text{Man(Jim), Woman(Jim)}\}$ consistent w.r.t. an empty TBox? **Yes**

- Is the above ABox $A$ consistent w.r.t. $T = \{\text{Woman} \equiv \text{Person} \land \text{Female}, \text{Man} \equiv \text{Person} \land \neg \text{Female}\}$? **No**
Automated reasoning

- State of the art DL systems typically use (highly optimised) tableaux algorithms

- Tableaux algorithms work by trying to construct a concrete example (model) consistent with KB axioms:
  - Start from ground facts (ABox axioms)
  - Explicate structure implied by complex concepts and TBox axioms
    - Syntactic decomposition using tableaux expansion rules
    - Infer constraints on (elements of) model
Tableaux Expansions rules

<table>
<thead>
<tr>
<th>Rule</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>□-rule</td>
<td>if 1. ((C_1 \cap C_2) \in \mathcal{L}(v)), (v) is not indirectly blocked, and 2. ({C_1, C_2} \not\subseteq \mathcal{L}(v)) then (\mathcal{L}(v) \rightarrow \mathcal{L}(v) \cup {C_1, C_2}).</td>
</tr>
<tr>
<td>⊖-rule</td>
<td>if 1. ((C_1 \cup C_2) \in \mathcal{L}(v)), (v) is not indirectly blocked, and 2. ({C_1, C_2} \cap \mathcal{L}(v) = \emptyset) then (\mathcal{L}(v) \rightarrow \mathcal{L}(v) \cup {E}) for some (E \in {C_1, C_2}).</td>
</tr>
<tr>
<td>∃-rule</td>
<td>if 1. (\exists r.C \in \mathcal{L}(v_1)), (v_1) is not blocked, and 2. (v_1) has no safe (r)-neighbour (v_2) with (C \in \mathcal{L}(v_1)), then create a new node (v_2) and an edge (\langle v_1, v_2 \rangle) with (\mathcal{L}(v_2) = {C}) and (\mathcal{L}(\langle v_1, v_2 \rangle) = {r}).</td>
</tr>
<tr>
<td>∀-rule</td>
<td>if 1. (\forall r.C \in \mathcal{L}(v_1)), (v_1) is not indirectly blocked, and 2. there is an (r)-neighbour (v_2) of (v_1) with (C \notin \mathcal{L}(v_2)) then (\mathcal{L}(v_2) \rightarrow \mathcal{L}(v_2) \cup {C}).</td>
</tr>
<tr>
<td>∀⁺-rule</td>
<td>if 1. (\forall r.C \in \mathcal{L}(v_1)), (v_1) is not indirectly blocked, and 2. there is some role (r') with (\text{Trans}(r')) and (r' \equiv r) 3. there is an (r')-neighbour (v_2) of (v_1) with (\forall r'.C \notin \mathcal{L}(v_2)) then (\mathcal{L}(v_2) \rightarrow \mathcal{L}(v_2) \cup {\forall r'.C}).</td>
</tr>
<tr>
<td>choose-rule</td>
<td>if 1. (\leq n r.C \in \mathcal{L}(v_1)), (v_1) is not indirectly blocked, and 2. there is an (r)-neighbour (v_2) of (v_1) with ({C, \neg C} \cap \mathcal{L}(v_2) = \emptyset) then (\mathcal{L}(v_2) \rightarrow \mathcal{L}(v_2) \cup {E}) for some (E \in {C, \neg C}).</td>
</tr>
<tr>
<td>⩾-rule</td>
<td>if 1. (\geq n r.C \in \mathcal{L}(v)), (v) is not blocked, and 2. there are not (n) safe (r)-neighbours (v_1, \ldots, v_n) of (v) with (C \in \mathcal{L}(v_i)) and (v_i \neq v_j) for (1 \leq i &lt; j \leq n).</td>
</tr>
</tbody>
</table>
DLs Reasoners

- Mature, highly optimized implementations
  - Research implementations
    - FaCT++, Pellet, KAON2, CEL, HermiT...
  - Commercial implementations
    - Cerebra, RacerPro, SHER...

(http://www.cs.manchester.ac.uk/~sattler/reasoners.html)
The Web Ontology Language (OWL) as a DL language

- OWL is W3C recommendation (i.e. a standard)
  - OWL DL is equivalent to the SHOIN
  - OWL Lite is equivalent to SHIF
  - Further connections issue from the recent OWL1.1 and OWL2 recommendations

- OWL exploits results of 15+ years of research in DLs
  - Well defined semantics
  - Complexity, decidability results
  - Reasoning algorithms
  - Implemented systems
## OWL class constructors

<table>
<thead>
<tr>
<th>Constructor</th>
<th>DL Syntax</th>
<th>Example</th>
<th>FOL Syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td>intersectionOf</td>
<td>$C_1 \sqcap \ldots \sqcap C_n$</td>
<td>Human $\sqcap$ Male</td>
<td>$C_1(x) \land \ldots \land C_n(x)$</td>
</tr>
<tr>
<td>unionOf</td>
<td>$C_1 \sqcup \ldots \sqcup C_n$</td>
<td>Doctor $\sqcup$ Lawyer</td>
<td>$C_1(x) \lor \ldots \lor C_n(x)$</td>
</tr>
<tr>
<td>complementOf</td>
<td>$\neg C$</td>
<td>$\neg$ Male</td>
<td>$\neg C(x)$</td>
</tr>
<tr>
<td>oneOf</td>
<td>${x_1} \sqcup \ldots \sqcup {x_n}$</td>
<td>${john} \sqcup {mary}$</td>
<td>$x = x_1 \lor \ldots \lor x = x_n$</td>
</tr>
<tr>
<td>allValuesFrom</td>
<td>$\forall P.C$</td>
<td>$\forall$hasChild.Doctor</td>
<td>$\forall y.P(x, y) \rightarrow C(y)$</td>
</tr>
<tr>
<td>someValuesFrom</td>
<td>$\exists P.C$</td>
<td>$\exists$hasChild.Lawyer</td>
<td>$\exists y.P(x, y) \land C(y)$</td>
</tr>
<tr>
<td>maxCardinality</td>
<td>$\leq n P$</td>
<td>$\leq$1hasChild</td>
<td>$\exists y.P(x, y)$</td>
</tr>
<tr>
<td>minCardinality</td>
<td>$\geq n P$</td>
<td>$\geq$2hasChild</td>
<td>$\exists y.P(x, y)$</td>
</tr>
</tbody>
</table>
## OWL axioms

<table>
<thead>
<tr>
<th>OWL Syntax</th>
<th>DL Syntax</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>subClassOf</td>
<td>$C_1 \sqsubseteq C_2$</td>
<td>Human $\sqsubseteq$ Animal $\sqcap$ Biped</td>
</tr>
<tr>
<td>equivalentClass</td>
<td>$C_1 \equiv C_2$</td>
<td>Man $\equiv$ Human $\sqcap$ Male</td>
</tr>
<tr>
<td>subPropertyOf</td>
<td>$P_1 \sqsubseteq P_2$</td>
<td>hasDaughter $\sqsubseteq$ hasChild</td>
</tr>
<tr>
<td>equivalentProperty</td>
<td>$P_1 \equiv P_2$</td>
<td>cost $\equiv$ price</td>
</tr>
<tr>
<td>transitiveProperty</td>
<td>$P^+ \sqsubseteq P$</td>
<td>ancestor$^+$ $\sqsubseteq$ ancestor</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OWL Syntax</th>
<th>DL Syntax</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>type</td>
<td>$a : C$</td>
<td>John : Happy-Father</td>
</tr>
<tr>
<td>property</td>
<td>$\langle a, b \rangle : R$</td>
<td>$\langle$John, Mary$\rangle$ : has-child</td>
</tr>
</tbody>
</table>
Summing up DLs

- Represent world (and its semantics) in terms of concepts, roles, individuals

- Very expressive formal knowledge representation languages with well-defined inference services

- Allow to handle effectively incomplete knowledge and reason over it
  - e.g., $\text{Tall} \sqcap \text{Intelligent} \ (\text{Tom}), \neg\text{Student} \ (\text{Jim})$

- Reasoning amounts to constructing logical models, hence the added complexity
  - Avoid using them to declare data structures and perform algorithmic computations
DLs in multimedia analysis related tasks

- DLs (ontologies) have been used for
  - Representation of domain specific knowledge for annotation purposes
  - Representation of media specific knowledge including low-level features and content structure/decomposition (so called multimedia ontologies)
  - Representation of domain-specific knowledge and linking with perceptual features in order to support the extraction of content semantics
Multimedia Ontologies (in a nutshell)

- Based in their majority on MPEG-7, multimedia ontologies aim to
  - Attach formal semantics to the XML-based Schema MPEG-7 definitions
  - Make explicit the normative specifications
  - Alleviate the ambiguities resulting from descriptions with multiple meanings

- Analysis, annotation, search, retrieval, presentation...
Multimedia Ontologies (cont.)

- Hunter’s ontology developed within the Harmony project (chronologically first initiative, 2001)

- The Multimedia Structure Ontology & Visual Descriptor Ontology, developed within aceMedia

- The Multimedia Content Ontology & Multimedia Descriptor Ontology, developed within BOEMIE

- SmartWeb, DS-MIRG, Rhizomik..

- The COMM (core) multimedia ontology (K-Space)
Hunter’s ontology (extract)
SmartWeb ontology (extract)
Boemie MCO ontology (extract)
Multimedia Ontologies

- Differences in
  - Coverage
  - Modeling
  - Level of axiomatisation (semantics clarity)
  - Linking with domain ontologies


S. Dasiopoulou, I. Kompatsiaris, M. G. Strintzis, **Enquiring MPEG-7 based Ontologies**, in Multimedia Tools and Appls., SI Data Semantics, 2009.
Knowledge-Based Semantics Extraction

- Huge topic, vast literature (dates back to 1970s, AI..)
  - focus on recent DL related approaches only
  - present (some) representative examples

- TBox: background knowledge
  - defines valid (coherent) interpretation

- ABox: extracted descriptions
  - analysis facts
The FUSION project (*)

Domain knowledge definition and linking with low-level representations
Scene Interpretation with DLs(*)

(*B.Neumann, R.Moller, On Scene Interpretation with Description Logics, in Cognitive Vision Systems, 2006.)
Scene interpretation with DLs (cont.)

- Scene interpretation as model construction
- Logical aggregates capture complex objects/events

(equivalent cover
  (and configuration
    (exactly 1 cv-pl plate)
    (exactly 1 cv-sc (and saucer (some near plate)))
    (exactly 1 cv-cp (and cup (some on saucer)))
    (subset cv-pl (compose cv-sc near))
    (subset cv-sc (compose cv-cp on))))

(image of a table setting with a plate, saucer, cup, and fork)
Scene interpretation with DLs (cont.)

- The available geometric descriptions are assumed to be
  - correct and non conflicting (which is not usually the case)
  - crisp (information loss)
  - complete (i.e. not missing)
Abductive reasoning for multimedia interpretation(*)

Abductive reasoning for multimedia interpretation (cont.)

- ABox abduction: given a background knowledge $\Sigma(T,A)$ and a set of observations $\Gamma$ derive a set of assertions $\Delta$ that \textbf{explain} $\Gamma$

  \[ \Sigma \cup \Delta \models \Gamma \]

  $\Gamma$ is divided into \textbf{bona fide} (that need to be explained) and \textbf{bona fide} (that are taken to be true) assertions
Abductive reasoning for multimedia interpretation (cont.)

Γ

bona fide

\[
\begin{align*}
& \text{Jumper} \sqsubseteq \text{Human} \\
& \text{Pole} \sqsubseteq \text{Sports\_Equipment} \\
& \text{Bar} \sqsubseteq \text{Sports\_Equipment} \\
& \text{Pole} \sqcap \text{Bar} \sqsubseteq \bot \\
& \text{Pole} \sqcap \text{Jumper} \sqsubseteq \bot \\
& \text{Jumper} \sqcap \text{Bar} \sqsubseteq \bot \\
& \text{Jumping\_Event} \sqsubseteq \exists_1 \text{hasParticipant, Jumper} \\
& \text{Pole\_Vault} \sqsubseteq \text{Jumping\_Event} \sqcap \exists \text{hasPart, Pole} \sqcap \exists \text{hasPart, Bar} \\
& \text{High\_Jump} \sqsubseteq \text{Jumping\_Event} \sqcap \exists \text{hasPart, Bar} \\
& \text{near}(Y, Z) \rightarrow \text{Pole\_Vault}(X), \text{hasPart}(X, Y), \text{Bar}(Y), \\
& \text{hasPart}(X, W), \text{Pole}(W), \text{hasParticipant}(X, Z), \text{Jumper}(Z) \\
& \text{near}(Y, Z) \rightarrow \text{High\_Jump}(X), \text{hasPart}(X, Y), \text{Bar}(Y), \\
& \text{hasParticipant}(X, Z), \text{Jumper}(Z)
\end{align*}
\]

\[
\begin{align*}
- \Delta_1 &= \{ \text{new\_ind}_1 : \text{Pole\_Vault}, (\text{new\_ind}_1, \text{bar}_1) : \text{hasPart}, (\text{new\_ind}_1, \text{new\_ind}_2) : \\
& \text{hasPart}, \text{new\_ind}_2 : \text{Pole}, (\text{new\_ind}_1, \text{human}_1) : \text{hasParticipant}, \text{human}_1 : \\
& \text{Jumper}\} \\
- \Delta_2 &= \{ \text{new\_ind}_1 : \text{Pole\_Vault}, (\text{new\_ind}_1, \text{bar}_1) : \text{hasPart}, (\text{new\_ind}_1, \text{pole}_1) : \\
& \text{hasPart}, \\
& (\text{new\_ind}_1, \text{human}_1) : \text{hasParticipant}, \text{human}_1 : \text{Jumper}\} \\
- \Delta_3 &= \{ \text{new\_ind}_1 : \text{High\_Jump}, (\text{new\_ind}_1, \text{bar}_1) : \text{hasPart}, (\text{new\_ind}_1, \text{human}_1) : \\
& \text{hasParticipant}, \\
& \text{human}_1 : \text{Jumper}\}
\end{align*}
\]
Enhancing Image Semantics Extraction using fuzzy DLs (*)

(S.Dasiopoulou, I.Kompatisiaris, M.G.Strintzis, Investigating fuzzy DLs-based Reasoning in Semantic Image Analysis, in Multimedia Tools and Apps., 2009.)
Uncertainty Issues

- Machine learning provides now generic methodologies for supporting more than 100 concepts
  - captures conveniently complex associations between perceptual features and semantics
  - successful application examples, yet variable general performance

- **Semantics goes beyond perceptual manifestations**
  - *possibly contradictory* (Mountain, Sand and Indoor)
  - *possibly overlapping / complementary* (Beach and Sea)
  - *of restricted abstraction* w.r.t. semantic expressiveness (face inside sea vs Swimmer)

- Learning-based extracted annotations need to be *semantically interpreted* into a *consistent* description
Semantics goes beyond perceptual manifestations

<table>
<thead>
<tr>
<th>Search Topic</th>
<th>Best Detector</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two visible tennis players on the court</td>
<td>Athlete</td>
<td>0.6501</td>
</tr>
<tr>
<td>A goal being made in a soccer match</td>
<td>Stadium</td>
<td>0.3429</td>
</tr>
<tr>
<td>Basketball players on the court</td>
<td>Indoor Sports Venue</td>
<td>0.2801</td>
</tr>
<tr>
<td>A meeting with a large table and people</td>
<td>Furniture</td>
<td>0.1045</td>
</tr>
<tr>
<td>People with banners or signs</td>
<td>People Marching</td>
<td>0.1013</td>
</tr>
<tr>
<td>One or more military vehicles</td>
<td>Armored Vehicles</td>
<td>0.0892</td>
</tr>
<tr>
<td>Helicopter in flight</td>
<td>Helicopters</td>
<td>0.0791</td>
</tr>
<tr>
<td>A road with one or more cars</td>
<td>Car</td>
<td>0.0728</td>
</tr>
<tr>
<td>An airplane taking off</td>
<td>Classroom</td>
<td>0.0526</td>
</tr>
<tr>
<td>A tall building</td>
<td>Office Building</td>
<td>0.0469</td>
</tr>
<tr>
<td>A ship or boat</td>
<td>Cloud</td>
<td>0.0427</td>
</tr>
<tr>
<td>George Bush entering or leaving vehicle</td>
<td>Rocket Propelled Grenades</td>
<td>0.0365</td>
</tr>
<tr>
<td>Omar Karami</td>
<td>Chair</td>
<td>0.0277</td>
</tr>
<tr>
<td>Graphic map of Iraq, Baghdad marked</td>
<td>Graphical Map</td>
<td>0.0269</td>
</tr>
<tr>
<td>Condoleeza Rice</td>
<td>US National Flag</td>
<td>0.0237</td>
</tr>
<tr>
<td>One or more palm trees</td>
<td>Weapons</td>
<td>0.0225</td>
</tr>
</tbody>
</table>

*Snoek et al., "Adding Semantics to Detectors for Video Retrieval", IEEE Multimedia, 2007*
Semantics goes beyond perceptual manifestations

- Conifers detector semantics pertain to mountainous scenes
- Sand detector semantics pertains to beach scenes

(image:ɔcontains.Sand) ≥ 0.75
(image:ɔcontains.Sky) ≥ 0.87
(image:ɔcontains.Foliage) ≥ 0.76
(image:ɔcontains.Conifers) ≥ 0.88
(image:ɛcontains.Landscape) ≥ 0.92

- Sea and Sand detectors entail Beach scene
- Beach scenes entails both Natural and Outdoor scenes

(image:ɔcontains.Sand) ≥ 0.75
(image:ɔcontains.Sea) ≥ 0.81
(image:ɔcontains.Person) ≥ 0.67
(image:ɔcontains.Foliage) ≥ 0.76
(image:ɔcontains.Grass) ≥ 0.58
(image:Beach) ≥ 0.85
(image:Beach) ≥ 0.67
Fuzzy DLs based approach

- **Goal:** enhance the robustness and completeness of learning-based extracted annotations

- **How:** semantics utilization
  - to interpret initial annotations
    - semantic integration
    - to detect and resolve inconsistencies
  - to enrich by means of entailment

- **Methodology:** fuzzy DL based reasoning
  - crisp TBox to conceptualize the domain semantics
  - fuzzy ABox to capture the uncertainty of initial annotations
Outdoor images TBox extract

Countryside_buildings ⊑ ∃contains.Buildings ⊍ ∃contains.Foliage
Countryside_buildings ⊑ Landscape
∃contains.Forest ⊍ ∃contains.Grass ⊍ ∃contains.Tree ⊑ ∃contains.Foliage
Rockyside ⊑ ∃contains.Cliff
Rockyside ⊑ ∃contains.Mountainous
Roadside ⊑ ∃contains.Road
Roadside ⊑ Landscape
∃contains.Sea ≡ Coastal
Coastal ⊏ Natural
∃contains.Forest ⊑ Landscape
Beach ≡ Coastal ⊍ ∃contains.Sand
Beach ⊏ Natural
Cityscape ⊑ ManMade
∃contains.Sky ⊑ Outdoor
∃contains.Trunk ⊑ ∃contains.Tree
Mountainous ⊍ Coastal ⊏ ⊥
Natural ⊍ ManMade ⊏ ⊥
Scene level interpretation

Domain TBox

Natural ≡ Outdoors ⊃ ManMade
Mountainous ≡ Natural ⊃ Coastal
Beach ≡ Coastal ⊃ ∃contains.Sand
∃contains.Mountain ⊆ Mountainous
∃contains.Sea ⊆ Coastal
∃contains.Sand ⊆ Mountainous ⊆ ⊥
Outdoor ⊆ Indoor ⊆ ⊥

Initial Assertions

(image:Indoor) ≥ 0.67
(image:∃contains.Sea) ≥ 0.73
(image:∃contains.Sand) ≥ 0.58
(image:∃contains.Mountain) ≥ 0.85

Disjointness axioms removed

Scene level hierarchy

Outdoor (0.85) ManMade
Natural (0.85)
Coastal (0.85) Mountainous (0.85)
Beach (0.58)
Indoor (0.67)
Consistency handling

Disjoint axioms restored

Domain TBox
Natural ≡ Outdoors ⊔ ~ ManMade
Mountainous ≡ Natural ⊔ ~ Coastal
Beach ≡ Coastal ⊓ ∃contains.Sand
∃contains.Mountain ⊓ Mountainous
∃contains.Sea ⊓ Coastal
∃contains.Sand ⊓ Mountainous ⊑ ⊥
Outdoor ⊓ Indoor ⊑ ⊥

Initial Assertions

(image:Indoor) ≥ 0.67
(image:∃contains.Sea) ≥ 0.73
(image:∃contains.Sand) ≥ 0.58
(image:∃contains.Mountain) ≥ 0.85

Scene level hierarchy

Outdoor (0.85) Indoor (0.67)
Natural (0.85) ManMade (0.65)
Coastal (0.58) Mountainous (0.85)
Beach (0.58)

Inconsistency handling

(image:∃contains.Mountain) ≥ 0.85
Enrichment

Initial Assertions

\[
\begin{align*}
(image: Indoor) & \geq 0.67 \\
(image: \exists contains. Sea) & \geq 0.73 \\
(image: \exists contains. Sand) & \geq 0.58 \\
(image: \exists contains. Mountain) & \geq 0.85
\end{align*}
\]

Disjointness axioms removed

\[
\begin{align*}
(image: Indoor) & \geq 0.67 \\
(image: \exists contains. Sea) & \geq 0.73 \\
(image: \exists contains. Sand) & \geq 0.58 \\
(image: Coastal) & \geq 0.73 \\
(image: Beach) & \geq 0.58 \\
(image: Natural) & \geq 0.73 \\
(image: Outdoor) & \geq 0.73 \\
(image: \exists contains. Mountain) & \geq 0.85 \\
(image: Mountainous) & \geq 0.85 \\
(image: Natural) & \geq 0.85 \\
(image: Outdoor) & \geq 0.85
\end{align*}
\]

Scene level hierarchy

T1 step

Domain TBox

\[
\begin{align*}
\text{Natural} & \equiv \text{Outdoors} \sqcup \neg \text{ManMade} \\
\text{Mountainous} & \equiv \text{Natural} \sqcup \neg \text{Coastal} \\
\text{Beach} & \equiv \text{Coastal} \sqcap \exists \text{contains. Sand} \\
\exists \text{contains. Mountain} & \sqsubseteq \text{Mountainous} \\
\exists \text{contains. Sea} & \sqsubseteq \text{Coastal} \\
\exists \text{contains. Sand} \sqcap \text{Mountainous} & \perp \\
\text{Outdoor} \sqcap \text{Indoor} & \perp
\end{align*}
\]

Disjoint axioms restored

T2 step

\[
\begin{align*}
(image: Indoor) & \geq 0.67 \\
(image: \exists contains. Sea) & \geq 0.73 \\
(image: \exists contains. Sand) & \geq 0.58 \\
(image: Coastal) & \geq 0.73 \\
(image: Beach) & \geq 0.58 \\
(image: \exists contains. Mountain) & \geq 0.85 \\
(image: Mountainous) & \geq 0.85 \\
(image: Natural) & \geq 0.85 \\
(image: Outdoor) & \geq 0.85
\end{align*}
\]

Inconsistency handling

\[
\begin{align*}
(image: \exists contains. Mountain) & \geq 0.85 \\
(image: Mountainous) & \geq 0.85 \\
(image: Natural) & \geq 0.85 \\
(image: Outdoor) & \geq 0.85
\end{align*}
\]

Final Assertions
Conclusions

- The use of explicit semantics is integral in multimedia semantics extractions; yet not the only necessary component.

- Handling uncertainty is a critical factor
  - formal handling of annotations uncertainty semantics
  - utilization of domain semantics
  - consistent interpretations / descriptions

- Largely misestimated degrees/analysis descriptions can mislead the interpretation.
Future Directions

- **Investigation of additional knowledge**
  - probabilistic information in the form of co-occurrence patterns
  - spatial relations among object level concepts (aligning different segmentation masks)

- **Investigation of intermediate representation level**
  - link domain definitions with qualitative visual features
    - inconsistent at domain level interpretations are not simply rejected

- **Experimentation with descriptions coming from other than image analysis sources**
  - text, tags (expressed in ontological terms)
    - provenance-based weights
Thank you for listening

Any Questions ??
Additional References

- C.Tsinaraki, P.Polydoros, S.Christodoulakis, Integration of OWL ontologies in MPEG-7 and TV Anytime Compliant Indexing, 16th Inter. Conf. on Advanced Information Systems Engineering (CAiSE), Riga, Latvia, 2004.