Interprétation d'images

Apport des ontologies et des logiques de description

Jamal Atif, Isabelle Bloch et Céline Hudelot

PSL, Université Paris Dauphine, CentreSupelec, Télécom-Paristech,
Université Paris Saclay

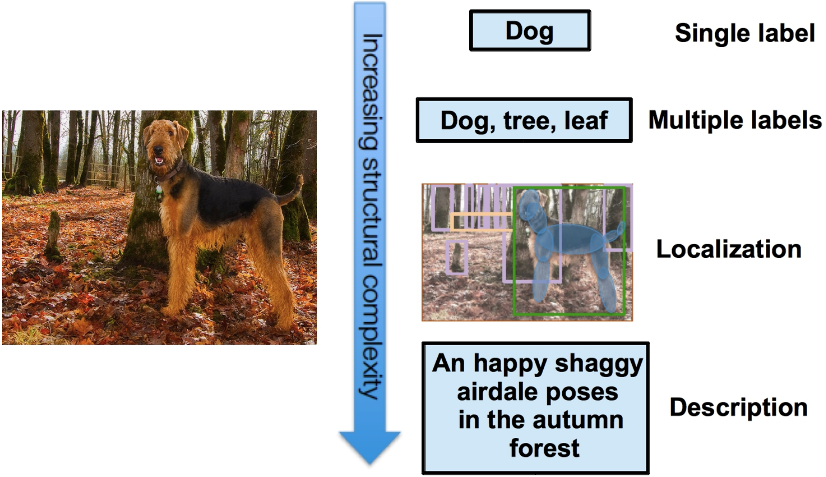
01 décembre 2015

Semantic image interpretation and annotation

Questions

What is the semantic content of these images? What do they represent?

Semantic image interpretation and annotation



Source : T Berg

Semantic image interpretation and annotation

A hard problem for machines in spite of the increasing performance of sensors and the computing capacities.

Issues [Smeulders 00, Snoek 10]

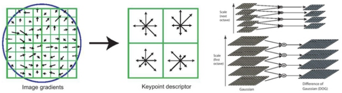
- Sensory gap.
- **Semantic gap**.
- Scaling gap: balance between expressivity/complexity and scaling of models.

Semantic image interpretation and annotation

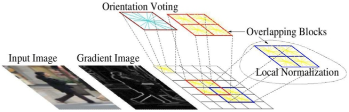
Sensory gap


Image = projection of a reality, often in 3D and continuous, into a discrete and 2D representation.

Numerous advances [Lowe 04, Dalal 05]



SIFT

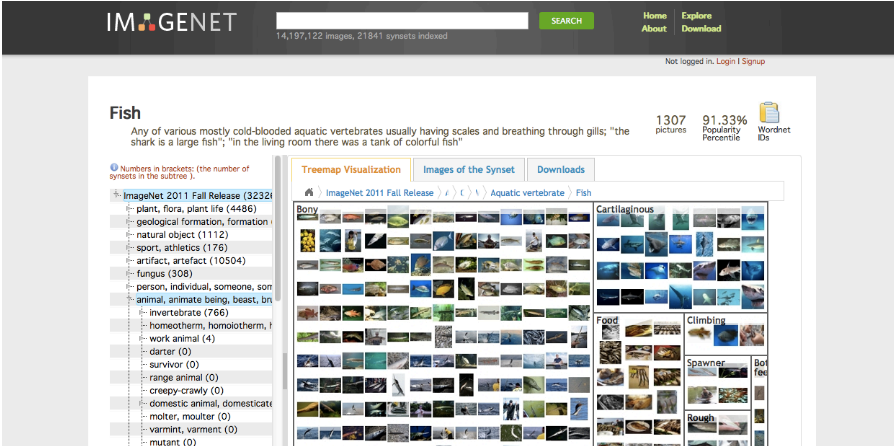


HoG

...

Semantic image interpretation and annotation

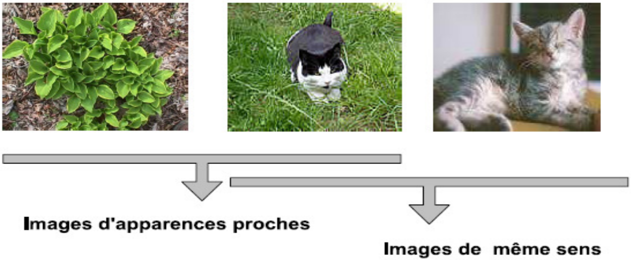
Scale gap



Convolutional Networks (Yann Le Cun) : [Krizhevsky 12, Erhan 14] : challenge ILSVRC : 1000 catgories et 1.461.406 images.

Semantic image interpretation and annotation

Semantic gap



Definition

Lack of coincidence between the information that one can extract from the visual data and the interpretation of these data by a user in a given situation [Smeulders 00].
Known as symbol grounding [Harnad 90] in AI and robotics.

Image and semantics

What is the semantics of this image?

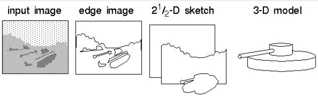
- *A white object on a green background.*
- *An insect.*
- *A white fly on a rose leaf.*

- Image semantics is not inside the image.
- Image interpretation depends on **a priori knowledge.**
- Image interpretation depends on the user objectives.
- Importance of contextual and structural information.

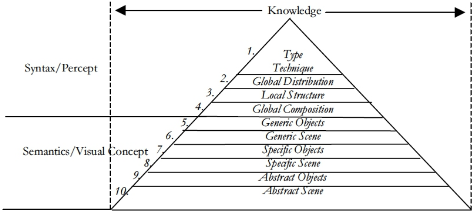
Image and semantics

A multi-level paradigm

Since the early years of CV



D. Marr hierarchy [Marr 82]



Semantic pyramid [Jaimes 00]

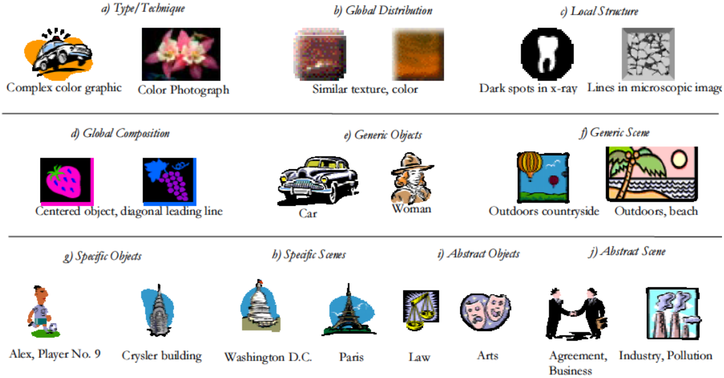
Niveau de la scène

Générique : Paysage de montage, rallye
Spécifique : Chypre
Abstrait : Sport, Divertissement

Niveau de l'objet

Générique : voiture, voiture de rallye
Spécifique : citroen de Sebastien Loeb

Image and semantics



Jaimes et al.

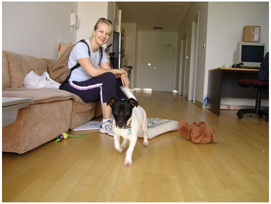
Image and semantics

Several semantics acceptations: from objects semantics to structural descriptions semantics.

→

Car: present
Cow: present
Bike: not present
Horse: not present
...

[Duygulu 02, Barnard 03, Lavrenko 03, Djeraba 03, Carneiro 07, Liu 07, Deng 10]



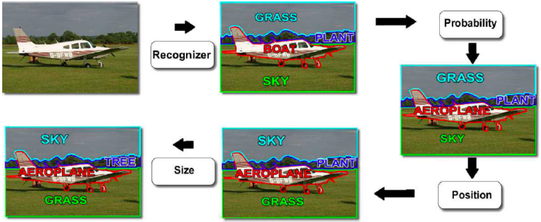
This is a photograph of one person and one brown sofa and one dog. The person is against the brown sofa. And the dog is near the person, and beside the brown sofa."

[Yao 10, Kulkarni 11, Farhadi 10, Farhadi 13, Karpathy 14]

Image and semantics

Importance of contextual and sptatial information

Source : [Parikh 12]



Source : [Galleguillos 10]

Importance of spatial relations in image interpretation

- Spatial reasoning
- Carry an important structural information
- More stable and reliable than object features

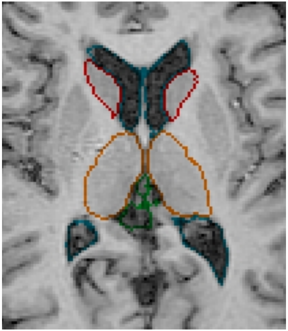
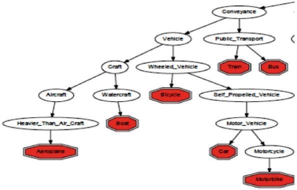


Image and semantics
Importance of prior knowledge

Semantics = a property that emerges from the interaction between data and knowledge [Hanson 78, Santini 01, Hudelot 03].

Connaissances implicites

Connaissances explicites



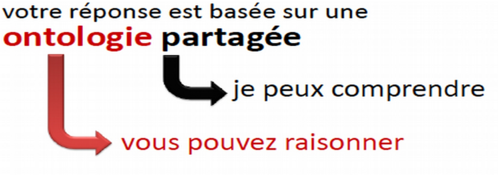
⇒ **Interest of ontologies**

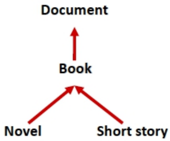
Ontology?

Source: F. Gandon, INRIA

What is the last document you have read?

Documents





Ontologies: definition

Ontology

Etymology : ontos (being, that which is) + logos (science, study, theory)

- Philosophy
  - Study of the nature of being or becoming
  - Study of the nature of existence or reality
- Informatique
  - Representing what exists in a formalism allowing for rational processing.
  - Explicit and formal specification of a given conceptualization [Gruber 95].

Ontologies : definition

ontology

Explicit and formal specification of a given conceptualization

- Explicit specification:
  - using a formal language (several languages are based on Description Logics).
- Conceptualization:
  - Structuring in concepts linked by relations.

Concept

- name
- meaning (definition in intension)
- denotation (definition in extension)

Relation

- name
- intension
- extension

Different types of ontologies

Depending on the abstraction level

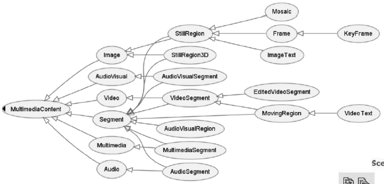
- **Top-ontology**: high-level knowledge with categories organized according to philosophical reflexions (ex: DOLCE).
- **Core-ontology**: basic and minimal ontology consisting only of the minimal concepts required to understand the other concepts (ex : Dublin Core).
- **Domain-ontology**: concepts and relations as manipulated by an expert of the domain (ex: FMA).

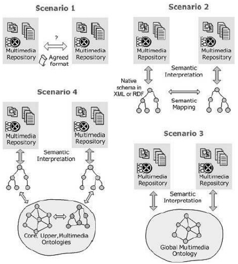
Ontologies for image annotation
Growing interest since 2001

Answering several issues:

- A unified and standardised description of concepts and primitives in the annotation process.
  - MPEG-7 ontologies [Hunter 01, Simou 05, Arndt 07, Dasiopoulou 10b].
  - Annotation vocabulary standardisation: LabelMe [Russell 08], ImageNet [Deng 09], LSCOM [Naphade 06], MediaMill [Snoek 06, Snoek 07].

Ontologies for standardised annotation vocabulary standardisation





Source : [Dasiopoulou 10b]

Ontologies for standardised annotation vocabulary standardisation

LSCOM LARGE SCALE CONCEPT ONTOLOGY FOR MULTIMEDIA

- natural hazard
  - earthquake
  - natural Disasters
  - tornado
  - avalanche
  - mudslide
- conveyance
  - airplane
    - flying
    - landing
    - take-off
  - ground vehicles
    - bus
    - truck
  - boat
    - sailboat
    - boat_ship

- ...
- computers
- recreational activity
  - sports
    - baseball
    - basketball
    - football
    - soccer
    - tennis
- group
- single person
  - single person male
  - single person female
- head and shoulders

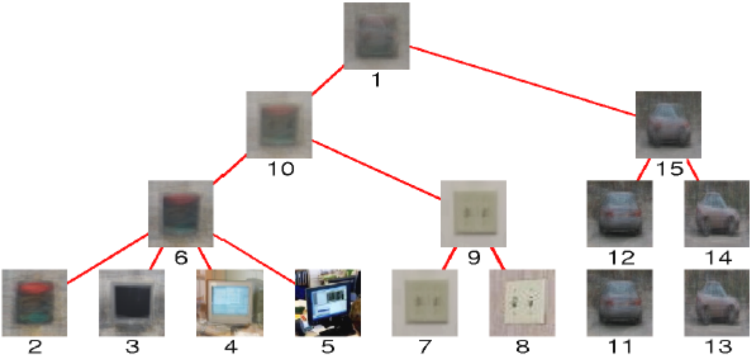
Ontologies for image annotation

Growing interest since 2001

Several types of hierarchies:

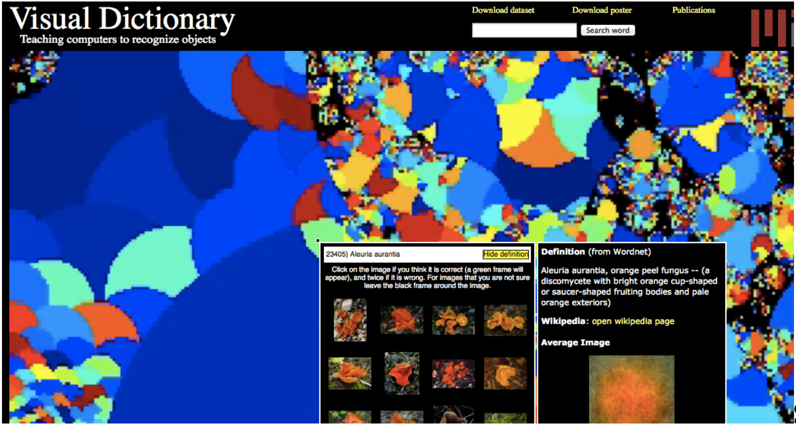
- Visual hierarchies:
  [Fei-Fei 05, Marszalek 08, Griffin 08, Sivic 08, Bart 08, Gao 11].
- Hierarchies constructed from lexical resources such as Wordnet:
  [Wei 07, Marszalek 07, Torralba 08].
- *Semantic* hierarchies: [Wu 12, Li 10, Fan 07, Fan 08, Shen 10]

Ontologies for image annotation



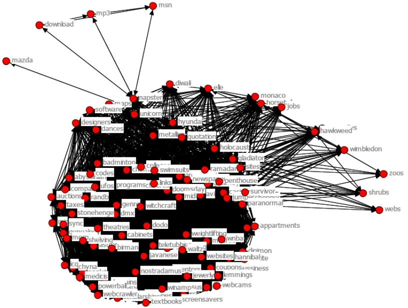
Visual hierarchies [Sivic 08].

Ontologies for image annotation



Source: http://groups.csail.mit.edu/vision/TinyImages/

Ontologies pour la structuration du vocabulaire d'annotation



Source: [Li 10, Wu 12]

Ontologies for image annotation

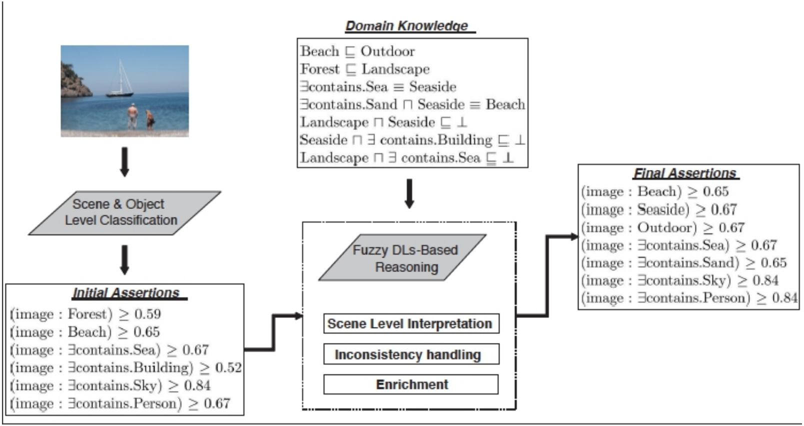
Based on lexical resources, e.g. wordnet

Figure : What are the most similar concepts?

According to Wordnet and a semantic similarity measure:

Distance(human, whale) = **7**
Distance(shark, whale) = **11**
Distance(human, shark) = **11**

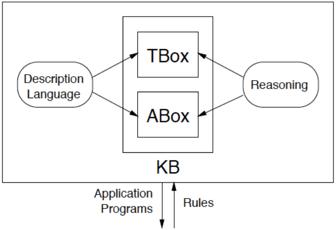
Ontologies for high-level image interpretation



Source : [Dasiopoulou 10a]

$\Rightarrow$ **Description Logics**

Description logics



A knowledge base (KB)

- TBox: introduces the *terminology*, i.e., the vocabulary of an application do- main
- ABox: contains *assertions* about named individuals in terms of this vocabulary.

Vocabulary: concepts, which denote sets of individuals, and roles, which denote binary relationships between individuals.

The basic description language $\mathcal{AL}$

Concept descriptions in $\mathcal{AL}$ are formed according to the following syntax rule:

$$
\begin{aligned}
C, D \longrightarrow & A \mid && \text{(atomic concepts)}\\
& \top \mid && \text{(universal concept)}\\
& \bot \mid && \text{(bottom concept)}\\
& \neg A \mid && \text{(atomic negation)}\\
& C \sqcap D \mid && \text{(intersection)}\\
& \forall R.C \mid && \text{(value restriction)}\\
& \exists.\top.C \mid && \text{(limited existential quantification).}
\end{aligned}
$$

The basic description language $\mathcal{AL}$

Example

- Atomic concepts: $Person, Female$
- Atomic role: *hasChlid*
- $\mathcal{AL}$-descriptions:

$$Person \sqcap Female$$
$$Person \sqcap \neg Female$$
$$Person \sqcap \exists hasChild.\top$$
$$Person \sqcap \forall hasChild.Female$$
$$Person \sqcap \forall hasChild.\bot$$

The basic description language $\mathcal{AL}$

Semantics

To give semantics to a logical system means to define a notion of truth for the formulas. In classical logic (dating back to Aristoteles) there are "only" two truth values "true" and "false" which we shall denote, respectively, by 1 and 0.

The basic description language $\mathcal{AL}$

Semantics

An interpretation $\mathcal{I} = \langle \Delta^{\mathcal{I}}, -^{\mathcal{I}} \rangle$

- $\Delta^{\mathcal{I}}$: a non-empty set, the domain of interpretation
- $-^{\mathcal{I}}$: an interpretation function, which assigns to every atomic concept A a set $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ and to every atomic role R a binary relation $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$.

Extension to concept descriptions:

$$\top^{\mathcal{I}} = \Delta^{\mathcal{I}}$$

$$\bot^{\mathcal{I}} = \emptyset$$

$$(\neg A)^{\mathcal{I}} = \Delta^{\mathcal{I}} \setminus A^{\mathcal{I}}$$

$$(A \sqcap D)^{\mathcal{I}} = C^{\mathcal{I}} \sqcap D^{\mathcal{I}}$$

$$(\forall R.C)^{\mathcal{I}} = \{a \in \Delta^{\mathcal{I}} \mid \forall b.(a,b) \in R^{\mathcal{I}} \rightarrow b \in C^{\mathcal{I}}\}$$

$$(\exists R.\bot)^{\mathcal{I}} = \{a \in \Delta^{\mathcal{I}} \mid \exists b.(a,b) \in R^{\mathcal{I}}\}$$

The basic description language $\mathcal{AL}$

Semantics

Equivalence:

$$C \equiv D \text{ if } C^{\mathcal{I}} = D^{\mathcal{I}} \text{ for all interpretations } \mathcal{I}$$

Example

$\forall hasChild.Female \sqcap \forall hasChild.Student$ and $\forall hasChild.(Female \sqcap Student)$ are equivalent.

The family of $\mathcal{AL}$ languages

$$\mathcal{AL}[\mathcal{U}][\mathcal{E}][\mathcal{N}][\mathcal{C}], \cdots$$

- Concept union ($\mathcal{U}$):

$$(C \sqcup D)^{\mathcal{I}} = C^{\mathcal{I}} \cup D^{\mathcal{I}}$$

- Full existential quantification ($\mathcal{E}$):

$$(\exists R.C)^{\mathcal{I}} = \{a \in \Delta^{\mathcal{I}} \mid \exists b.(a, b) \in R^{\mathcal{I}} \text{ and } b \in C^{\mathcal{I}}\}$$

- Number restriction ($\mathcal{N}$):

$$(\geq nR)^{\mathcal{I}} = \{a \in \Delta^{\mathcal{I}} \mid |\{b \mid (a, b) \in R^{\mathcal{I}}\}| \geq n\},$$
$$(\leq nR)^{\mathcal{I}} = \{a \in \Delta^{\mathcal{I}} \mid |\{b \mid (a, b) \in R^{\mathcal{I}}\}| \leq n\},$$

- Arbitrary concept negation ($\mathcal{C}$):

$$(\neg C)^{\mathcal{I}} = \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$$

The family of $\mathcal{AL}$ languages

$\mathcal{ALEN}$ example

$$Person \sqcap (\leq 1\ hasChild \sqcup (\geq 3\ hasChild \sqcap \exists hasChild.Female))$$

Terminologies

- Terminological axioms

$$C \sqsubseteq D \text{ (general concept inclusions)}, \; C \equiv D \text{ (general concept equalities)}$$
$$R \sqsubseteq S \text{ (general role inclusions)}, \; R \equiv S \text{ (general role equalities)}$$

- An interpretation $\mathcal{I}$ satisfies an inclusion $C \sqsubseteq D$ if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$

$$\mathcal{I} \models (C \sqsubseteq D) \Leftrightarrow C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$$

- An interpretation $\mathcal{I}$ satisfies an equality $C \equiv D$ if $C^{\mathcal{I}} \equiv D^{\mathcal{I}}$

$$\mathcal{I} \models (C \equiv D) \Leftrightarrow C^{\mathcal{I}} \equiv D^{\mathcal{I}}$$

Definitions

- Introduce *symbolic names* for complex descriptions

$$Mother \equiv Woman \sqcap \exists hasChild.Person$$

- A finite set of definitions is called a *terminology* or TBox and is denoted $\mathcal{T}$

$$\begin{aligned}
Woman &\equiv Person \sqcap Female \\
Man &\equiv Person \sqcap \neg Woman \\
Mother &\equiv Woman \sqcap \exists hasChild.Person \\
Father &\equiv Man \sqcap \exists hasChild.Person \\
Parent &\equiv Father \sqcup Mother \\
Grandmother &\equiv Mother \sqcap \exists hasChild.Parent \\
MotherWithManyChildren &\equiv Mother \sqcap \geq 3 hasChild
\end{aligned}$$

Terminologies

- $\mathcal{I}$ is a model of a TBox $\mathcal{T}$ if it satisfies all GCIs in $\mathcal{T}$

$$\mathcal{I} \models \mathcal{T}$$

- Two TBoxes are equvalent if they have the same model.

Assertional knowledge, ABox

- Describes a specific state of affairs of an application domain in terms of concepts and roles

Let C be a concept and R a role, an assertion is of the form

- $C(a)$: concept assertion
- $R(a, b)$: role assertion

a, b are called individual names

- Given $\mathcal{I}$, each individual a is mapped to an element $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$
- Unique name assumption: $a^{\mathcal{I}} \neq b^{\mathcal{I}}$

Assertional knowledge, ABox

- An ABox is a finite set of assertions.

$\mathcal{I}$ is a model of the ABox $\mathcal{A}$ if it satisfies all its assertions:

- $a^{\mathcal{I}} \in C^{\mathcal{I}}$ for all $C(a) \in \mathcal{A}$
- $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in R$ if for all $R(a, b) \in \mathcal{A}$

A model $\mathcal{I}$ satisfies an assertion α or an ABox $\mathcal{A}$ with respect to a TBox $\mathcal{T}$ if in addition to being a model of α or of $\mathcal{A}$, it is a model of $\mathcal{T}$.

$$\mathcal{I} \models_{\mathcal{T}} \alpha \iff \mathcal{I} \models \mathcal{T} \text{ and } \mathcal{I} \models \alpha$$

Example

MotherWithoutDaughter(MARY)
hasChild(MARY,PETER)
hasChild(MARY,PAUL)

Father(PETER)
hasChild(PETER, HARRY)

An example of a knowledge base and its model

$$\mathcal{KB} = \{\texttt{MusicLover} \sqsubseteq \exists\texttt{hasFriend.Musician},$$
$$\texttt{hasFriend}(\texttt{Peter}, \texttt{Paul}),$$
$$(\forall\texttt{hasFriend.}\neg\texttt{Musician})(\texttt{Peter}),$$
$$\texttt{MusicLover}(\texttt{Paul}),$$
$$\texttt{Peter} \not\approx \texttt{Paul}\}$$

$$\Delta = \{\texttt{Peter}, \texttt{Paul}, \texttt{x}\}$$
$$\texttt{Peter}^{\mathcal{I}} = \texttt{Peter}$$
$$\texttt{Paul}^{\mathcal{I}} = \texttt{Paul}$$
$$\texttt{x}^{\mathcal{I}} = \texttt{x}$$
$$\texttt{Musician}^{\mathcal{I}} = \{\texttt{x}\}$$
$$\texttt{MusicLover}^{\mathcal{I}} = \{\texttt{Paul}\}$$
$$\texttt{hasFriend}^{\mathcal{I}} = \{(\texttt{Peter}, \texttt{Paul}), (\texttt{Paul}, \texttt{x})\}$$

Concrete domains

- A way to integrate *concrete and quantitative qualities* (intergers, strings,...) of real world objects with conceptual knowledge [Baader,91].
- A pair (Δ_D, Φ_D) where Δ_D is a set and Φ_D a set of predicates names on Δ_D. Each predicate name P is associated with an arity n and an n-ary predicate $P^{\mathcal{D}} \subseteq \Delta_D^n$

Examples

- Concrete domain $\mathcal{N}$
  - domain : non negative integers
  - predicates : $\leq$ (binary predicate) $\leq n$ unary predicate
  - $\mathsf{Person} \sqcap \exists \mathsf{age}. \leq 20$ denotes a person whose age is less than 20
- Concrete domain $\mathcal{AL}$: Allen's interval calculus
  - domain : intervals
  - predicates : built from Allen's basic interval relations

Ontologies, Knowledge bases (KB)

An ontology or a KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ is composed of a TBox $\mathcal{T}$ and an ABox $\mathcal{A}$

An interpretation $\mathcal{I}$ is a model of a KB if it is a model of $\mathcal{T}$ and of $\mathcal{A}$

$$\mathcal{I} \models \mathcal{K} \iff \forall \alpha \in \mathcal{T} \cup \mathcal{A}, \mathcal{I} \models \alpha$$

Reasoning services

$\Longrightarrow$ Infer implicit knowledge from explicitly one

- Terminological reasoning
- Assertional reasoning

Terminological reasoning

Satisfiability: C is satisfiable w.r.t. a TBox $\mathcal{T}$ iff $C^{\mathcal{I}} \neq \emptyset$ for some model $\mathcal{I}$ of $\mathcal{T}$

Subsumption: C is subsumed by D w.r.t. a TBox $\mathcal{T}$ ($C \sqsubseteq_{\mathcal{T}} D$) iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ for all models $\mathcal{I}$ of $\mathcal{T}$

Equivalence: C is equivalent to D w.r.t. a TBox $\mathcal{T}$ ($C \equiv_{\mathcal{T}} D$) iff $C^{\mathcal{I}} = D^{\mathcal{I}}$ for all models $\mathcal{I}$ of $\mathcal{T}$

Disjointness: Two concepts C and D are disjoint with respect to $\mathcal{T}$ if $C^{\mathcal{I}} \cap D^{\mathcal{I}} = \emptyset$ for every model $\mathcal{I}$ of $\mathcal{T}$

$\sqsubseteq_{\mathcal{T}}$ is a pre-order (reflexive and transitive)

Reduction to subsumption

For concepts C, D we have

- C is unsatisfiable $\iff$ C is subsumed by $\bot$;
- C and D are equivalent $\iff$ C is subsumed by D and D is subsumed by C;
- C and D are disjoint $\iff$ $C \cup D$ is subsumed by $\bot$.

The statements also hold with respect to a TBox.

Reduction to Unsatisfiability

For concepts C, D we have

- C is subsumed by $D \iff C \sqcap \neg D$ is unsatisfiable;
- C and D are equivalent $\iff$ both $C \sqcap \neg D$ and $\neg C \sqcap D$ are satisfiable;
- C and D are disjoint $\iff C \sqcap D$ is unsatisfiable.

The statements also hold with respect to a TBox.

Reducing Unsatisfiability

Let C be a concept. Then the following are equivalent:

- C is unsatisfiable;
- C is subsumed by $\bot$;
- C and $\bot$ are equivalent;
- C and $\bot$ are disjoint.

The statements also hold with respect to a TBox.

Assertional reasoning

Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be an ontology.

Consistency $\mathcal{A}$ is consistent with respect to a TBox $\mathcal{T}$, if there is an interpretation that is a model of both $\mathcal{A}$ and $\mathcal{T}$

Instance checking a is an instance of C w.r.t. $\mathcal{T}$ iff $a^{\mathcal{I}} \in C^{\mathcal{I}}$ for all models $\mathcal{I}$ of $\mathcal{T}$. We also write $\mathcal{A} \models C(a)$. The same holds for roles.

Retrieval problem Given an ABox $\mathcal{A}$ and a concept C, find all individuals a such that $\mathcal{A} \models C(a)$.

Realization problem (dual to the retrieval problem). Given an individual a and a set o concepts, find *the most specific concepts* (msc) C from the set such that $A \models C(a)$. The mscs are the concepts that are minimal with respect to the subsumption ordering $\sqsubseteq$.

Reduction

- $\mathcal{A} \models C(a)$ iff $\mathcal{A} \cup \{\neg C(a)\}$ is inconsistent;
- C is satisfiable iff $\{C(a)\}$ is consistent.

Subsumption checking

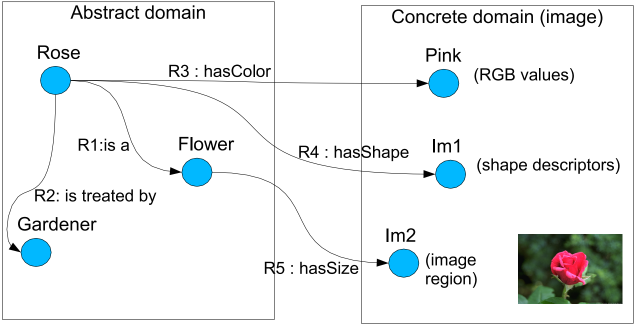
- Structural subsumption
- Semantic tableaux
- etc.

A spatial relation ontology for semantic image interpretation

Hudelot et al. [Hudelot 08]

Ontologies, concrete domains and semantic gap

Hudelot et al. [Hudelot 08]



Idea

Each application domain concept is linked to its representation in the image domain.

Importance of spatial relations

Hudelot et al. [Hudelot 08]

Spatial reasoning

Largely developed in the artificial intelligence community

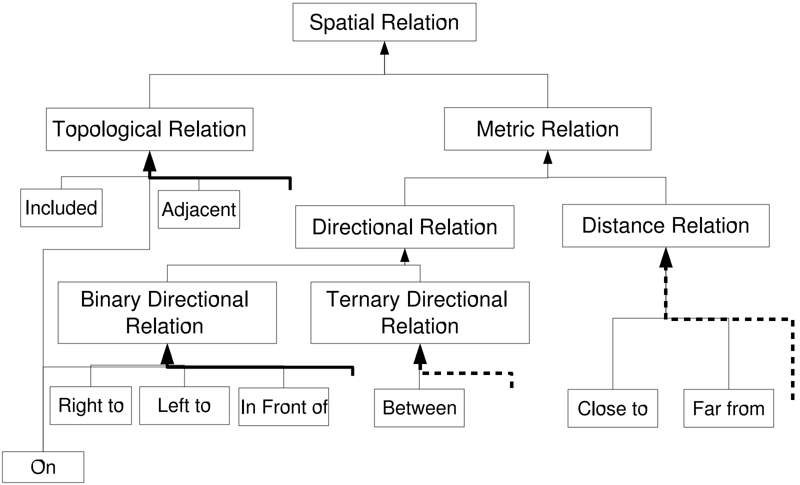
- Mainly topological relations
- Formal logics (ex: mereotopology)
- Inference

Less developed in image interpretation

- Need for imprecise knowledge representation
- (Semi-)quantitative framework ($\Rightarrow$ numerical evaluation)
- Examples: structural recognition in images under imprecision

A spatial relation ontology

Hudelot et al. [Hudelot 08]



Reference system

Hudelot et al. [Hudelot 08]

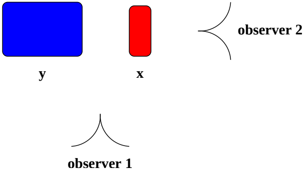
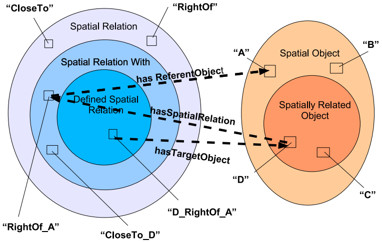


Figure : Directional relation **Right Of**

- The relation is dependent on the viewer.
- Three concepts are necessary to define a spatial relation:
  - Target object;
  - Reference object;
  - Reference system (ex: relative vs absolute).

Formal representation of spatial relations

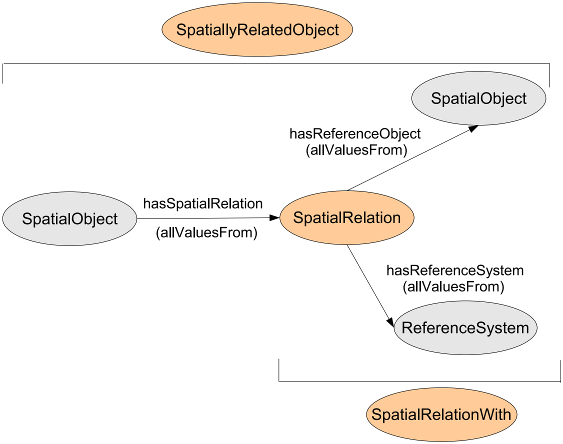


The nature of spatial relations is twofold:

- Concepts with their own properties
- Links between concepts

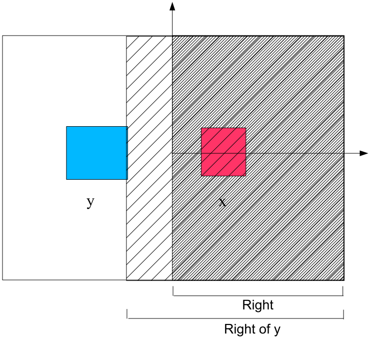
Formal representation of spatial relations

Hudelot et al. [Hudelot 08]



Formal representation of spatial relations

Hudelot et al. [Hudelot 08]



x is to the right of y: true

Abox:

- y:SpatialObject; x:SpatialObject
- Right_Of_y $\equiv$ Right_Of $\sqcap$ $\ni$hasReferentObject.{y}
- x:SpatialObject $\sqcap$ $\exists$ hasSpatialRelation.Right_Of_y and x:SpatiallyRelatedObject
- C_0 $\equiv$ SpatialRelation $\sqcap$ $\ni$hasReferentObject.{y} $\sqcap$ $\ni$hasTargetObject.{x}

Importance of fuzzy representation

Hudelot et al. [Hudelot 08]

- Representation of imprecision:
  - objects (no clear boundaries, coarse segmentation...)
  - relations (ex: *left of, quite close*)
  - type of knowledge available (ex: *the caudate nucleus is close to the lateral ventricle*)
  - question to be answered (ex: *go towards this object while remaining at some security distance*)
- Two classes of relations:
  - well defined in the crisp case (adjacency, distances...)
  - vague even in the crisp case (directional relationships...)
- Fusion of several and heterogeneous pieces of knowledge and information

Fuzzy representations of spatial relations

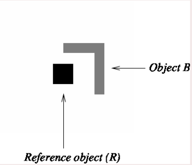
Hudelot et al. [Hudelot 08]

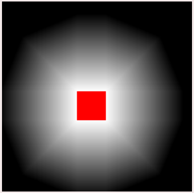
The representation depends on:

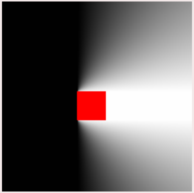
- The class of the relation
- The type of reasoning

Two questions:

- Given two objects (possibly fuzzy), assess the degree to which a relation is satisfied
- Given one reference object, define the area of the space in which a relation to this reference is satisfied (to some degree)

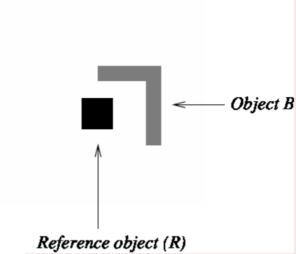


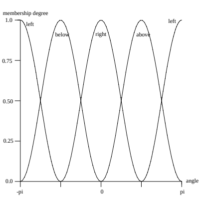




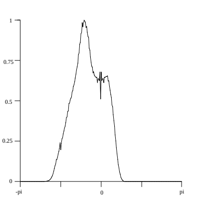
Fuzzy representation of directional relations

Hudelot et al. [Hudelot 08]

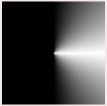




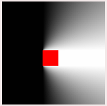
Fuzzy sets



Angle Histogram



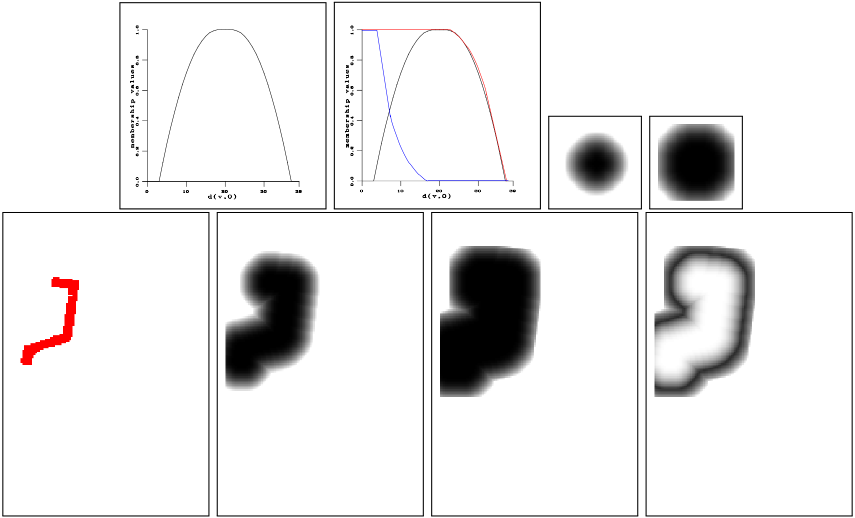
Right Of



Right Of R

Fuzzy representation of distance relations

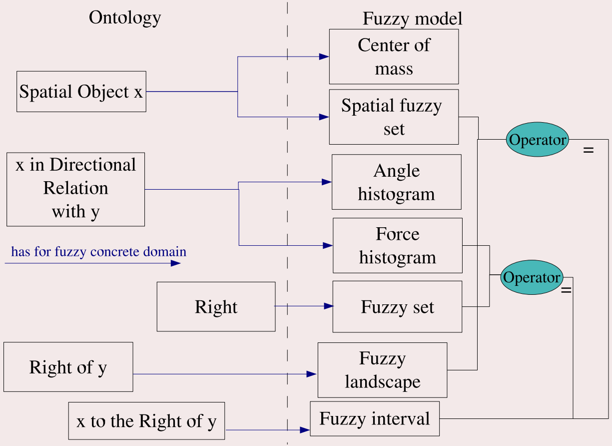
Hudelot et al. [Hudelot 08]



Ontology and fuzzy model integration

Hudelot et al. [Hudelot 08]

Fuzzy concrete domains[Nagypal,03]



Knowledge in brain imaging

Hudelot et al. [Hudelot 08]

Concepts:

- brain: part of the central nervous system located in the head
- caudate nucleus: a deep gray nucleus of the telencephalon involved with control of voluntary movement
- glioma: tumor of the central nervous system that arises from glial cells
- ...

Spatial organization:

- the left caudate nucleus is inside the left hemisphere
- it is close to the lateral ventricle
- it is outside (left of) the left lateral ventricle
- it is above the thalamus, etc.
- ...

Description of anatomical knowledge

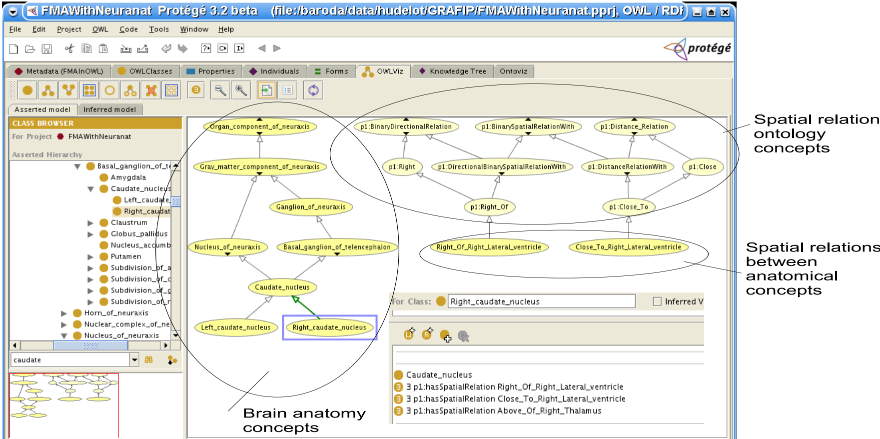
Hudelot et al. [Hudelot 08]

Tbox:

- $\mathsf{AnatomicalStructure} \sqsubseteq \mathsf{SpatialObject}$
- $\mathsf{GN} \sqsubseteq \mathsf{AnatomicalStructure}$
- $\mathsf{LV} \equiv \mathsf{RLV} \sqcup \mathsf{LLV}$
- $\mathsf{LV} \equiv \mathsf{RLV} \sqcup \mathsf{LLV}$
- $\mathsf{CN} \doteq \mathsf{GN} \sqcap \exists\, \mathsf{hasSR.(Close\_To\_LV)}$
- $\mathsf{CN} \equiv \mathsf{RCN} \sqcup \mathsf{LCN}$
- etc.

Knowledge Representation

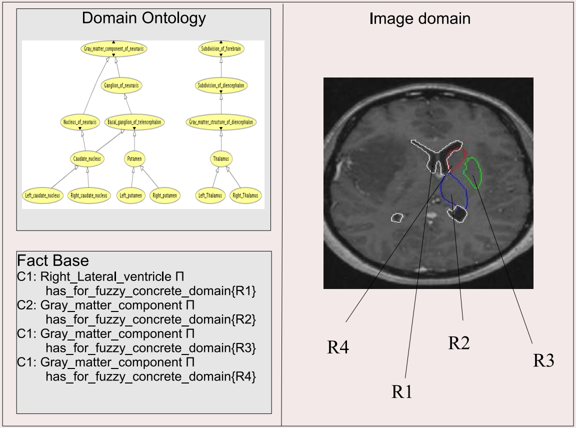
Hudelot et al. [Hudelot 08]



First question: global approach

Hudelot et al. [Hudelot 08]

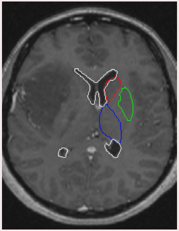
Several objects are first extracted from the images using a segmentation method and then recognized.

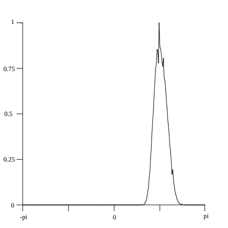


Question 1: global approach

Hudelot et al. [Hudelot 08]

Several objects are first extracted from the images using a segmentation methods and then recognized.

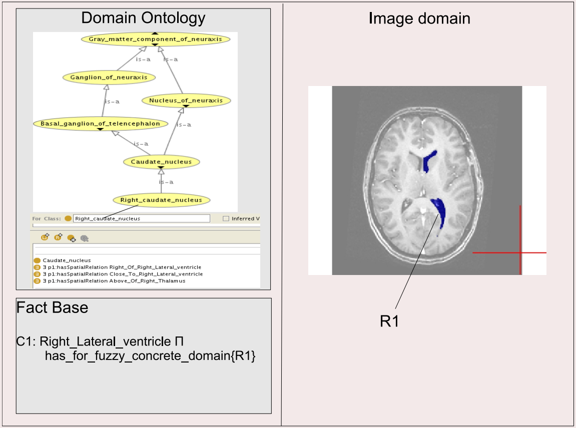




Question 2: sequential approach

Hudelot et al. [Hudelot 08]

The structure are recognized successively

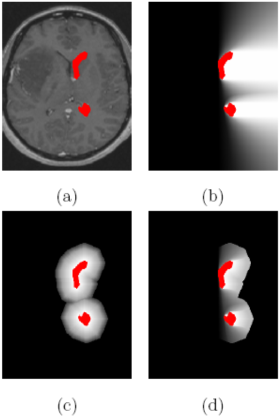


Question 2: sequential approach

Hudelot et al. [Hudelot 08]

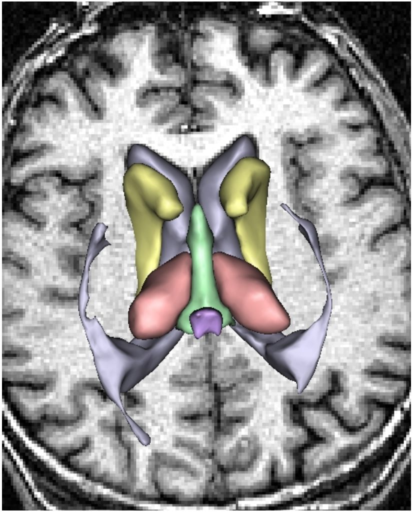
The Caudate Nucleus is:

- To the right of the Lateral Ventricle
- Close to the Lateral Ventricle



Results: healthy case

Hudelot et al. [Hudelot 08]



Non-monotonic reasoning for image interpetation

- Default reasoning
- Abductive reasoning

Image interpretation as a default reasoning service

$$\frac{\alpha : \beta_1, \cdots, \beta_n}{\gamma}$$

- α: precondition of the rule
- β_i: justifications
- γ: consequent

Intuitive explanation

Starting with a world description α of what is known to be true, i.e. deducible and it is consistent to assume β_i then conclude γ.

Example :
$\forall x, plays\_instruments(x) : improvises(x) / jazz\_musician(x)$
For all x is x plays an instrument and if the fact that x can improvise is consistent with all other knowledge then we can conclude that x is a jazz musician.

Default reasoning in DL

A terminological default theory: $(\mathcal{A}, \mathcal{D})$

- $\mathcal{A}$: ABox
- $\mathcal{D}$: finite set of terminological rules whose preconditions, justifications and consequents are concept terms.

Maintaining decidability

- Closed default rules: α, β_i, γ are ABox concept axioms (no use of free variables, i.e. TBox concept axioms).

Spatioterminological default reasoning.

Moller et al. approach [Möller 99, Neumann 08]

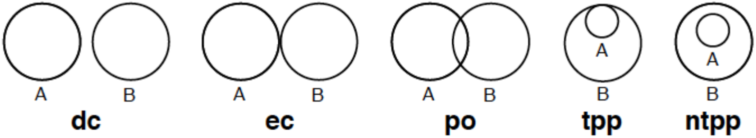
$\mathcal{ALCRP}(\mathcal{S}_2)$ Aboxes inside the default rules

$\mathcal{ALCRP}(\mathcal{S}_2)$

ALC with *predicate existence restriction* and a concrete domain $\mathcal{S}_2$ defined w.r.t. the topological space $\langle \mathbb{R}^2, 2^{\mathbb{R}^2} \rangle$

Spatioterminological default reasoning.

Moller et al. approach [Möller 99, Neumann 08]



The concrete domain $\mathcal{S}_2 = \langle \mathbb{R}^2, 2^{\mathbb{R}^2} \rangle$

- Predicate $is - region$ with $is - region^{\mathcal{S}_2} = \Delta_{\mathcal{S}_2}$ and its negation $is - no - region$ with $is - no - region^{\mathcal{S}_2} = 0_{\mathcal{S}_2}$
- 8 basic predicates $dc, ec, po, tpp, ntpp, tppi, eq$
- Predicates to name disjunctions of base relations
- The predicate $dc - ec - po - tpp - ntpp - tppi - ntppi - eq$ is called $spatially - related$
- A binary predicate $inconsistent - relation$ with $inconsistent - relation^{\mathcal{S}_2} = \emptyset$ is the negation of $spatially - related$

Spatioterminological default reasoning.

Moller et al. approach [Möller 99, Neumann 08]

Example 1

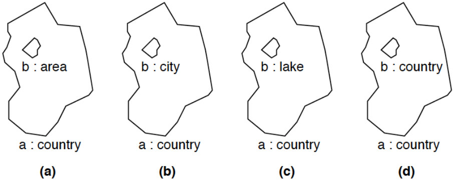


Figure : Interpretation pb: generate hypotheses for object b.

S_2 predicates formalization

$$
\begin{aligned}
inside &\doteq \exists(has\_area)(has\_area).\text{tpp-ntpp} \\
contains &\doteq \exists(has\_area)(has\_area).\text{tppi-ntppi} \\
overlaps &\doteq \exists(has\_area)(has\_area).\text{po} \\
touches &\doteq \exists(has\_area)(has\_area).\text{ec} \\
disjoint &\doteq \exists(has\_area)(has\_area).\text{dc}
\end{aligned}
$$

Spatioterminological default reasoning.

Moller et al. approach [Möller 99, Neumann 08]

Example 1

TBox

$$\begin{aligned}
area &\doteq \exists(has_area).\text{is-region}\\
natural_region &\doteq \neg administrative_region\\
country_region &\dot\sqsubseteq administrative_region \sqcap\\
&\quad large_scale \sqcap area\\
city_region &\dot\sqsubseteq administrative_region \sqcap\\
&\quad \neg large_scale \sqcap area\\
lake_region &\dot\sqsubseteq natural_region \sqcap area\\
river_region &\dot\sqsubseteq natural_region \sqcap area
\end{aligned}$$

$$\begin{aligned}
country &\doteq country_region \sqcap\\
&\quad \forall contains.\neg country_region \sqcap\\
&\quad \forall overlaps.\neg country_region \sqcap\\
&\quad \forall inside.\neg country_region\\
city &\doteq city_region \sqcap\\
&\quad \exists inside.country_region\\
lake &\dot\sqsubseteq lake_region\\
river &\doteq river_region \sqcap\\
&\quad \forall overlaps.\neg lake_region \sqcap\\
&\quad \forall contains.\bot \sqcap\\
&\quad \forall inside.\neg lake_region
\end{aligned}$$

Spatioterminological default reasoning.

Moller et al. approach [Möller 99, Neumann 08]

Example 1

Abox

$\{a : country, b : area, (a,b) : contains, (b,a) : inside\}$

Spatioterminological default rules

$$d_1 = \frac{area : city}{city} \quad d_2 = \frac{area : lake}{lake} \quad d_3 = \frac{area : city}{city}$$

Closed spatioterminological default rules, $d_i(ind)$

e.g.

$$d_1(a) = \frac{\{a : area\} : \{a : city\}}{\{a : city\}}$$

Spatioterminological default reasoning.

Moller et al. approach [Möller 99, Neumann 08]

Example 1

Default rules reasoning

- $d_1(a)$: cannot be applied. A contradiction between $a : city$ and $a : country$ in the Abox. *country_region* and *city_region* are disjoint in the TBox (due to *large_scale* and ¬*large_scale*).
- $d_1(b)$: ca be applied. Abox extension:

$$\{a : country, b : area, b : city, (a, b) : contains, (b, a) : inside\}$$

- $d_2(a)$: cannot be applied. A contradiction between $a : lake$ and $a : country$ in the Abox. *administrative_region* and *natural_region* are disjoint.
- $d_2(b)$: can be applied. Abox extension:

$$\{a : country, b : area, b : lake, (a, b) : contains, (b, a) : inside\}$$

But if Abox contains $d_1(a)$, $d_2(b)$ cannot be applied $\Longrightarrow$ two possible extensions.

Spatioterminological default reasoning.

Moller et al. approach [Möller 99, Neumann 08]

Example 1

Default rules reasoning, cont'd

- $d_3(a)$ cannot be applied. Its conclusion is already entailed by the ABox.
- $d_3(b)$ cannot be applied. The consequent $b : country$ makes the Abox inconsistent because a is already known as a country.

$$\mathcal{A} \models (a : \forall contains.\neg country\_region)$$
$$(a, b) : contains, b : country \implies b : country\_region$$

Spatioterminological default reasoning.

Moller et al. approach [Möller 99, Neumann 08]

Example 2

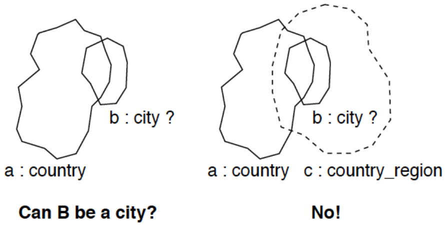


Figure : Subtle inferences due to topological constraints

Abox

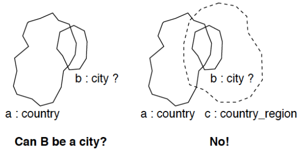
$$\{a : country, b : area, (a, b) : overlaps, (b, a) : overlaps\}$$

$\Longrightarrow$ the default rule $d_1(b)$ cannot be applied to conclude that object b is a city.

Spatioterminological default reasoning.

Moller et al. approach [Möller 99, Neumann 08]

Example 2



$$\mathcal{A} = \{a : country, b : area, (a, b) : overlaps, (b, a) : overlaps\}$$

$(b, a) : overlaps, b : city \implies b : city\_region \sqcap \exists inside.country\_region \implies \not\models (a : country\_region)$ (since $(b, a) : overlaps$).

Remark

Due to $\exists$ there exists an implicit individual c which is a *country_region* such that $(b, c) : inside$ hold which is impossible due to topological constraints (b inside c and c not overlap with a or does not contain a).

Abductive reasoning

- Abduction using safe rules (Peraldi et al. [Peraldi 07])
- Concept abduction (Atif et al. [Atif 14])

Abduction using safe rules

Peraldi et al. [Peraldi 07]

- Multimedia interpretation as abduction problem
- Use of conjunctive queries:

$$\{(X_1, \cdots, X_n) \mid atom_1, \cdots, atom_m\}, \text{ with}$$
$$atom = C(X), R(X, Y), (X = Y)$$

Example :
$\{x \mid \exists y \exists z (ChildOf(x, y) \wedge ChildOf(x, z) \wedge Married(y, z))\}$

Formalisation

Peraldi et al. [Peraldi 07]

Given an ABox assertions Γ in a form of a query, and a KB $\Sigma = (\mathcal{T}, \mathcal{A})$ derive all sets of Abox assertions Δ (explanations) such that $\Sigma \cup \Delta = \Gamma$ and the following conditions are satisfied:

- $\Sigma \cup \Delta$ is satisfiable, and
- Δ is a minimal explanation for Γ, i.e. there exists no other explanation Δ' in the solution set that is not equivalent to Δ and it holds that $\Sigma \cup \Delta' \models \Delta$.

Formalisation

Peraldi et al. [Peraldi 07]

Multimedia abduction:

$$\Sigma \cup \Gamma_1 \cup \Delta \models \Gamma_2$$

- $\Sigma = (\mathcal{T}, \mathcal{A})$, with $\mathcal{A}$ assumed empty
- $\Gamma = \Gamma_1 \cup \Gamma_2$, set of Abox assertions, encoding extracted objects from images and their spatial relationships
- Γ_1: bona fide assertions, assumed to be true by default
- Γ_2: assertions requiring fiats (aimed to be explained)
- Δ: a set of ABox explanations.

The process is implemented as (boolean) query answering.

Peraldi et al. [Peraldi 07]

Fig. 1. A pole vault event.

$$\begin{aligned} pole_1 &: Pole \\ human_1 &: Human \\ bar_1 &: Bar \\ (bar_1, human_1) &: near \end{aligned}$$

Fig. 2. An Abox Γ representing the results of low-level image analysis.

$$\begin{aligned}
Jumper &\sqsubseteq Human \\
Pole &\sqsubseteq Sports_Equipment \\
Bar &\sqsubseteq Sports_Equipment \\
Pole \sqcap Bar &\sqsubseteq \bot \\
Pole \sqcap Jumper &\sqsubseteq \bot \\
Jumper \sqcap Bar &\sqsubseteq \bot \\
Jumping_Event &\sqsubseteq \exists_{\leq 1} hasParticipant.Jumper \\
Pole_Vault &\sqsubseteq Jumping_Event \sqcap \exists hasPart.Pole \sqcap \exists hasPart.Bar \\
High_Jump &\sqsubseteq Jumping_Event \sqcap \exists hasPart.Bar \\
near(Y,Z) &\leftarrow Pole_Vault(X), hasPart(X,Y), Bar(Y), \\
&\quad\; hasPart(X,W), Pole(W), hasParticipant(X,Z), Jumper(Z) \\
near(Y,Z) &\leftarrow High_Jump(X), hasPart(X,Y), Bar(Y), \\
&\quad\; hasParticipant(X,Z), Jumper(Z)
\end{aligned}$$

Fig. 3. A tiny example Σ consisting of a Tbox and DL-safe rules.

The approach by Espinoda et al.

Illustration

Fig. 1. A pole vault event.

$$\begin{aligned} pole_1 &: Pole \\ human_1 &: Human \\ bar_1 &: Bar \\ (bar_1, human_1) &: near \end{aligned}$$

Fig. 2. An Abox Γ representing the results of low-level image analysis.

- $\Gamma_1 = \{pole_1 : Pole, human_1 : Human, bar_1 : Bar\}$
- $\Gamma_2 = \{(bar_1, human_1) : near\}$
- Boolean query $Q_1 := \{() \mid near(bar_1, human_1)\}$

Peraldi et al. [Peraldi 07]

- $\Delta_1 = \{new\_ind_1 : Pole\_Vault, (new\_ind_1, bar_1) : hasPart, (new\_ind_1, new\_ind_2) : hasPart, new\_ind_2 : Pole, (new\_ind_1, human_1) : hasParticipant, human_1 : Jumper\}$
- $\Delta_2 = \{new\_ind_1 : Pole\_Vault, (new\_ind_1, bar_1) : hasPart, (new\_ind_1, pole_1) : hasPart, new\_ind_1 : human_1, hasParticipant, human_1 : Jumpe\}$
- $\Delta_3 = \{new\_ind_1 : HighJump, (new\_ind_1, bar_1) : hasPart, (new\_ind_1, human1) : hasParticipant, human_1 : Jumper\}$

Preference score :

$$S_p(\Delta) := S_i(\Delta) - S_h(\Delta), \text{ with}$$
$$S_i(\Delta) := |\{i \mid i \in inds(\Delta) \text{ and } i \in inds(\Sigma \cup \Gamma_1)\}|$$
$$S_h(\Delta) := |\{i \mid i \in inds(\Delta) \text{ and } i \in new\_inds\}|$$

Peraldi et al. [Peraldi 07]

- Δ_1 incorporates $human_1$ and bar_1 from Γ_1, then $S_i(\Delta_1) = 2$
- Δ_1 hypothesizes two new individuals: new_ind_1, new_ind_2, then $S_h(\Delta_1) = 2$

$\Longrightarrow$ $S_p(\Delta_1) = 0$

- $S_p(\Delta_2) = 2$
- $S_p(\Delta_3) = 1$

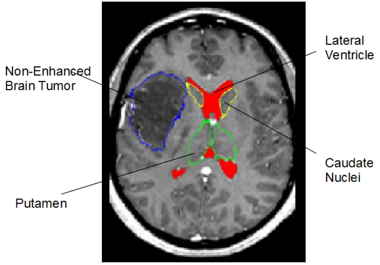
$\Longrightarrow$ Δ_2 represents the 'preferred' explanation:

$$\Delta_2 = \{new\_ind_1 : Pole\_Vault, (new\_ind_1, bar_1) : hasPart, (new\_ind_1, pole_1) : hasPart, new\_ind_1 : human_1, hasParticipant, human_1 : Jumper\}$$

Brain image understanding

Atif et al. [Atif 14]

image interpretation



Pathological brain with small deforming peripheral tumor

Interpretation as an abduction process

$\mathcal{K} \models (\gamma \rightarrow \varphi)$

Computing of the *best explanation* from observations φ given some a priori expert knowledge encoded in description logics

Knowledge representation

$$
\begin{aligned}
\textit{CerebralHemisphere} &\sqsubseteq \textit{BrainAnatomicalStructure} \\
\textit{PeripheralCerebralHemisphere} &\sqsubseteq \textit{CerebralHemisphereArea} \\
\textit{SubCorticalCerebralHemisphere} &\sqsubseteq \textit{CerebralHemisphereArea} \\
\textit{GreyNuclei} &\sqsubseteq \textit{BrainAnatomicalStructure} \\
\textit{LateralVentricle} &\sqsubseteq \textit{BrainAnatomicalStructure} \\
\textit{BrainTumor} &\sqsubseteq \textit{Disease} \\
&\quad \sqcap \exists \textit{hasLocation}.\textit{Brain} \\
\textit{SmallDeformingTumor} &\equiv \textit{BrainTumor} \\
&\quad \sqcap \exists \textit{hasBehavior}.\textit{Infiltrating} \\
&\quad \sqcap \exists \textit{hasEnhancement}.\textit{NonEnhanced} \\
\textit{SubCorticalSmallDeformingTumor} &\equiv \textit{SmallDeformingTumor} \sqcap \\
&\quad \exists \textit{hasLocation}.\textit{SubCorticalCerebralHemisphere} \\
&\quad \sqcap \exists \textit{closeTo}.\textit{GreyNuclei} \\
\textit{PeripheralSmallDeformingTumor} &\equiv \textit{BrainTumor} \sqcap \\
&\quad \exists \textit{hasLocation}.\textit{PeripheralCerebralHemisphere} \\
&\quad \sqcap \exists \textit{farFrom}.\textit{LateralVentricle}
\end{aligned}
$$

$$
\begin{aligned}
\textit{LargeDefTumor} &\equiv \textit{BrainTumor} \sqcap \\
&\quad \exists \textit{hasLocation}.\textit{CerebralHem} \\
&\quad \sqcap \exists \textit{hasComponent}.\textit{Edema} \\
&\quad \sqcap \exists \textit{hasComponent}.\textit{Necrosis} \\
&\quad \sqcap \exists \textit{hasEnhancement}.\textit{Enhanced} \\
&\ldots
\end{aligned}
$$

Initial ABox $\mathcal{A}_1$

$\{t_1 : \textit{BrainTumor};\ e_1 : \textit{NonEnhanced};\ l_1 : \textit{LateralVentricle};\ p_1 : \textit{PeripheralCerebralHemisphere};\ (t_1, e_1) : \textit{hasEnhancement};\ (t_1, l_1) : \textit{farFrom};\ (t_1, p_1) : \textit{hasLocation};\ \ldots\}$.

Interpretation as a concept abduction process

$\mathcal{K} \models \gamma \sqsubseteq O$, with O defined as

$$BrainTumor \sqcap \exists hasEnhancement.NonEnhanced \sqcap \exists farFrom.LateralVentricle \sqcap \exists hasLocation.PeripheralCerebralHemisphere$$

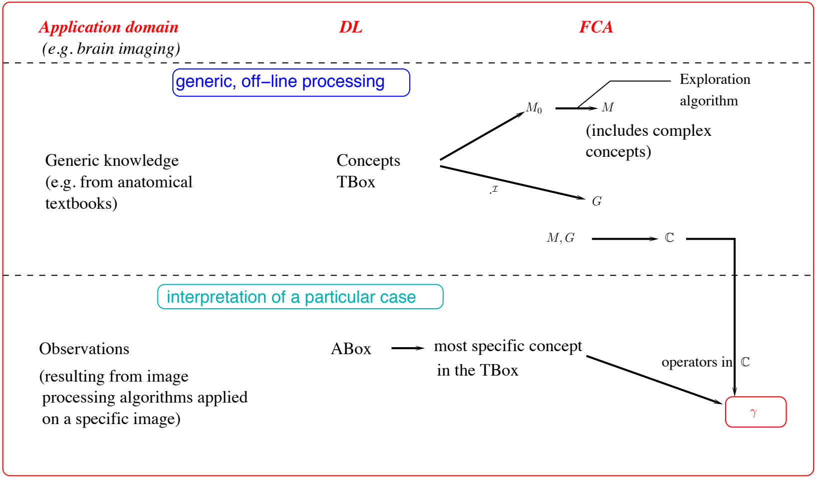
A set of possible explanations is :

$$\{DiseasedBrain, SmallDeformingTumoralBrain, PeripheralSmallDeformingTumoralBrain\}$$

The preferred solution according to minimality constraints is:

$$\gamma \equiv PeripheralSmallDeformingTumoralBrain$$

Global scheme



$\mathbb{K}_{brain}$

$\mathbb{K}_{brain}$	HumanOrgan	Brain	CerebralHemisphere	BrainAnatomicalStructure	CerebralHemisphereArea	PeripheralCerebralHemisphereArea	SubCorticalCerebralHemisphereArea	GreyNuclei	LateralVentricle	Disease	Edema	Necrosis	Enhancing	NonEnhancing	Infiltrating	hasLocation_Brain	BrainTumor	...
b_1	X	X														X		
b_2	X	X														X		
b_3	X	X														X		
b_4	X	X														X		
b_5	X	X														X		
b_6	X	X														X		
b_7	X	X														X		
gn_1				X				X										
gn_2				X				X										
gn_3				X				X										
gn_4				X				X										
lv_1				X					X									
lv_2				X					X									
t_1										X						X	X	
t_2										X						X	X	
t_3										X						X	X	

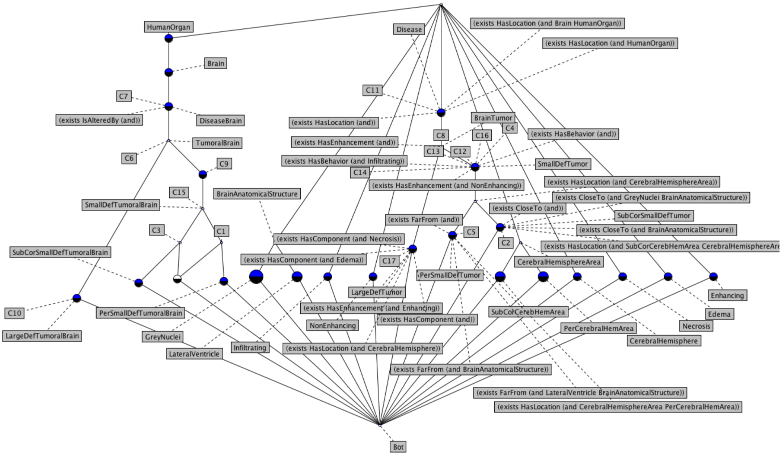


Figure : Excerpt of the concept lattice induced by the formal context $\mathbb{K}_{brain}$.

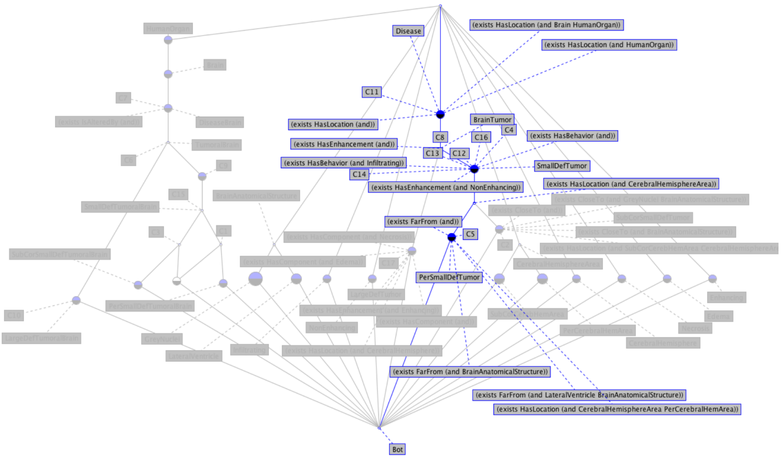


Figure : The erosion path leading to compute the preferred explanation.

- Richard Arndt, Raphaël Troncy, Steffen Staab, Lynda Hardman & Miroslav Vacura.
  COMM: designing a well-founded multimedia ontology for the web.
  In Proceedings of the 6th international The semantic web and 2nd Asian conference on Asian semantic web conference, ISWC'07/ASWC'07, pages 30–43, Berlin, Heidelberg, 2007. Springer-Verlag.

- J. Atif, C. Hudelot & I. Bloch.
  Explanatory reasoning for image understanding using formal concept analysis and description logics.
  IEEE Transactions on Systems, Man and Cybernetics: Systems, vol. 44, no. 5, pages 552–570, May 2014.

- Kobus Barnard, Pinar Duygulu, David Forsyth, Nando de Freitas, David M. Blei & Michael I. Jordan.
  Matching words and pictures.

Journal of Machine Learning Research, vol. 3, pages 1107–1135, 2003.

Evgeniy Bart, Ian Porteous, Pietro Perona & Max Welling.
Unsupervised learning of visual taxonomies.
In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08), 2008.

Gustavo Carneiro, Antoni B. Chan, Pedro J. Moreno & Nuno Vasconcelos.
Supervised Learning of Semantic Classes for Image Annotation and Retrieval.
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 3, 2007.

Navneet Dalal & Bill Triggs.
Histograms of Oriented Gradients for Human Detection.

In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 20-26 June 2005, San Diego, CA, USA, pages 886–893, 2005.

Stamatia Dasiopoulou & Ioannis Kompatsiaris.
Trends and Issues in Description Logics Frameworks for Image Interpretation.
Artificial Intelligence: Theories, Models and Applications, vol. 6040, pages 61–70, 2010.

Stamatia Dasiopoulou, Vassilis Tzouvaras, Ioannis Kompatsiaris & Michael G. Strintzis.
Enquiring MPEG-7 based multimedia ontologies.
Multimedia Tools and Applications (MTAP'10), vol. 46, pages 331–370, 2010.

J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li & L. Fei-Fei.
ImageNet: A Large-Scale Hierarchical Image Database.

In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'09), 2009.

Jia Deng, Alexander C. Berg, Kai Li & Li Fei-Fei.
What does classifying more than 10,000 image categories tell us?
In Proceedings of the European Conference on Computer Vision (ECCV'10), 2010.

Chabane Djeraba.
Association and Content-Based Retrieval.
IEEE Transaction on Knowledge and Data Engineering (TDDE'03), vol. 15, pages 118–135, January 2003.

P. Duygulu, Kobus Barnard, J. F. G. de Freitas & David A. Forsyth.
Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary.
In Proceedings of the European Conference on Computer Vision (ECCV'02), pages 97–112. Springer-Verlag, 2002.

Dumitru Erhan, Christian Szegedy, Alexander Toshev & Dragomir Anguelov.
Scalable Object Detection Using Deep Neural Networks.
In 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014, pages 2155–2162, 2014.

Jianping Fan, Yuli Gao & Hangzai Luo.
Hierarchical classification for automatic image annotation.
In international ACM SIGIR conference on Research and development in information retrieval (SIGIR'07), pages 111–118, 2007.

Jianping Fan, Yuli Gao & Hangzai Luo.
Integrating Concept Ontology and Multitask Learning to Achieve More Effective Classifier Training for Multilevel Image Annotation.
IEEE Transactions on Image Processing (TIP'08), vol. 17, no. 3, pages 407–426, 2008.

- Ali Farhadi, Mohsen Hejrati, Mohammad Amin Sadeghi, Peter Young, Cyrus Rashtchian, Julia Hockenmaier & David Forsyth.
  Every Picture Tells a Story: Generating Sentences from Images.
  In Proceedings of the 11th European Conference on Computer Vision: Part IV, ECCV'10, pages 15–29, Berlin, Heidelberg, 2010. Springer-Verlag.

- Ali Farhadi & Mohammad Amin Sadeghi.
  Phrasal Recognition.
  IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 12, pages 2854–2865, December 2013.

- L. Fei-Fei & P. Perona.
  A Bayesian hierarchical model for learning natural scene categories.

In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2005), volume 2, pages 524 – 531, 2005.

Carolina Galleguillos & Serge J. Belongie.
Context based object categorization: A critical survey.
Computer Vision and Image Understanding, vol. 114, no. 6, pages 712–722, 2010.

Tianshi Gao & Daphne Koller.
Discriminative learning of relaxed hierarchy for large-scale visual recognition.
In Proceedings of the International Conference on Computer Vision (ICCV'11), 2011.

Gregory Griffin & Pietro Perona.
Learning and Using Taxonomies for Fast Visual Categorization.
In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08), 2008.

Thomas R. Gruber.
Toward principles for the design of ontologies used for knowledge sharing.
International Journal of Human-Computer Studies, vol. 43, no. 5-6, pages 907–928, 1995.

Allen R Hanson & Edward M Riseman.
VISIONS: A computer system for interpreting scenes.
Computer vision systems, vol. 78, 1978.

Stevan Harnad.
The symbol grounding problem.
Physica D: Nonlinear Phenomena, vol. 42, no. 1, pages 335–346, 1990.

C. Hudelot & M. Thonnat.
A cognitive vision platform for automatic recognition of natural complex objects.

15th IEEE International Conference on Tools with Artificial Intelligence, 2003. Proceedings., pages 398–405, 2003.

C. Hudelot, J. Atif & I. Bloch.
Fuzzy Spatial Relation Ontology for Image Interpretation.
Fuzzy Sets and Systems, vol. 159, no. 15, pages 1929–1951, 2008.

Jane Hunter.
Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology.
In Semantic Web Working Symposium (SWWS'01), pages 261–281, 2001.

Alejandro Jaimes & Shih fu Chang.
A Conceptual Framework for Indexing Visual Information at Multiple Levels.
In Storage and Retrieval for Image and Video Databases (SPIE'00), pages 2–15, 2000.

Andrej Karpathy, Armand Joulin & Li Fei-Fei.
Deep Fragment Embeddings for Bidirectional Image Sentence Mapping.
CoRR, vol. abs/1406.5679, 2014.

Alex Krizhevsky, Ilya Sutskever & Geoffrey E. Hinton.
ImageNet Classification with Deep Convolutional Neural Networks.
In Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States., pages 1106–1114, 2012.

G. Kulkarni, V. Premraj, S. Dhar, Siming Li, Yejin Choi, A.C. Berg & T.L. Berg.
Baby talk: Understanding and generating simple image descriptions.

In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pages 1601 –1608, 2011.

V. Lavrenko, R. Manmatha & J. Jeon.
A model for learning the semantics of pictures.
In Neural Information Processing Systems (NIPS'03). MIT Press, 2003.

Li-Jia Li, Chong Wang, Yongwhan Lim, David M. Blei & Fei-Fei Li.
Building and using a semantivisual image hierarchy.
In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'10), pages 3336 –3343, 2010.

Ying Liu, Dengsheng Zhang, Guojun Lu & Wei-Ying Ma.
A survey of content-based image retrieval with high-level semantics.
Pattern Recognition, vol. 40, no. 1, pages 262 – 282, 2007.

David G. Lowe.
Distinctive Image Features from Scale-Invariant Keypoints.
International Journal of Computer Vision, vol. 60, no. 2, pages 91–110, 2004.

David Marr.
Vision: A computational investigation into the human representation and processing of visual information.
W. H. Freeman, 1982.

Marcin Marszalek & Cordelia Schmid.
Semantic hierarchies for visual object recognition.
In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07), pages 1 –7, 2007.

Marcin Marszalek & Cordelia Schmid.
Constructing Category Hierarchies for Visual Recognition.
In Proceedings of the European Conference on Computer Vision (ECCV), pages 479–491, 2008.

Ralf Möller, Bernd Neumann & Michael Wessel.
Towards computer vision with description logics: Some recent progress.
In spelmg, page 101. IEEE, 1999.

Milind Naphade, John R. Smith, Jelena Tesic, Shih-Fu Chang, Winston Hsu, Lyndon Kennedy, Alexander Hauptmann & Jon Curtis.
Large-Scale Concept Ontology for Multimedia.
IEEE MultiMedia, vol. 13, pages 86–91, July 2006.

Bernd Neumann & Ralf Möller.
On scene interpretation with description logics.
Image and Vision Computing, vol. 26, no. 1, pages 82–101, 2008.

Devi Parikh, C. Lawrence Zitnick & Tsuhan Chen.
Exploring Tiny Images: The Roles of Appearance and Contextual Information for Machine and Human Object Recognition.

IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 10, pages 1978–1991, October 2012.

S Espinosa Peraldi, A Kaya, S Melzer, R Möller & M Wessel.
Multimedia interpretation as abduction.
In Proc. dl-2007: International workshop on description logics. Citeseer, 2007.

Bryan C. Russell, Antonio Torralba, Kevin P. Murphy & William T. Freeman.
LabelMe: A Database and Web-Based Tool for Image Annotation.
International Journal of Computer Vision, vol. 77, no. 1-3, pages 157–173, 2008.

Simone Santini, Amarnath Gupta & Ramesh Jain.
Emergent Semantics through Interaction in Image Databases.

IEEE Transactions on Knowledge and Data Engineering, vol. 13, no. 3, pages 337–351, 2001.

Yi Shen & Jianping Fan.
Leveraging loosely-tagged images and inter-object correlations for tag recommendation.
In Proceedings of the international conference on Multimedia (MM'10), pages 5–14, 2010.

N. Simou, V. Tzouvaras, Y. Avrithis, G. Stamou & S. Kollias.
A Visual Descriptor Ontology for Multimedia Reasoning.
In Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS'05), 2005.

J. Sivic, B. C. Russell, A. Zisserman, W. T. Freeman & A. A. Efros.
Unsupervised Discovery of Visual Object Class Hierarchies.

In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08), 2008.

Arnold W. M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta & Ramesh Jain.
Content-based image retrieval at the end of the early years.
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, pages 1349–1380, 2000.

Cees G. M. Snoek, Marcel Worring, Jan C. van Gemert, Jan-Mark Geusebroek & Arnold W. M. Smeulders.
The challenge problem for automated detection of 101 semantic concepts in multimedia.
In Proceedings of the 14th annual ACM international conference on Multimedia, ACM MM'06, pages 421–430, 2006.

C. G.M. Snoek, B. Huurnink, L. Hollink, M. de Rijke, G. Schreiber & M. Worring.
Adding Semantics to Detectors for Video Retrieval.

IEEE Transactions on Multimedia (TMM'07), vol. 9, no. 5, pages 975–986, 2007.

- Cees G. M. Snoek & Arnold W. M. Smeulders.
  Visual-Concept Search Solved?
  IEEE Computer, vol. 43, no. 6, pages 76–78, June 2010.

- A. Torralba, R. Fergus & W.T. Freeman.
  80 Million Tiny Images: A Large Data Set for Nonparametric Object and Scene Recognition.
  Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 30, no. 11, pages 1958 –1970, 2008.

- Xiao-Yong Wei & Chong-Wah Ngo.
  Ontology-enriched semantic space for video search.
  In International Conference on Multimedia (MM'07), pages 981–990, 2007.

- Lei Wu, Xian-Sheng Hua, Nenghai Yu, Wei-Ying Ma & Shipeng Li.

Flickr Distance: A Relationship Measure for Visual Concepts.
Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 34, no. 5, pages 863 –875, 2012.

B.Z. Yao, Xiong Yang, Liang Lin, Mun Wai Lee & Song-Chun Zhu.
I2T: Image Parsing to Text Description.
Proceedings of the IEEE, vol. 98, no. 8, pages 1485 –1508, 2010.