Semantics and Multimedia

Motivation

Giuseppe Serra
http://www.micc.unifi.it/serra
serra@dsi.unifi.it
Semantics: What for?

• Tasks and purpose
  - Find pictures about a person in mountain
  - Find a cute picture about Irene to show to my friends
  - Find pictures from Civetta Mountain
  - Find pictures from our common holiday

other restrictions:
  - with a resolution of …
  - with on other person in the foreground

All of this information can not be retrieved using Google engine.
What does this mean?

Usually to find these pictures in your laptop you must browse all your pictures in a chronologic timeline.
Multimedia Semantics

- Content semantics (general knowledge)
  - person, snow
- Content semantics (private knowledge)
  - Irene
- Retrieval semantics:
  - “cute” picture to show
- Situation semantics:
  - Civetta, Dolomiti 2007
- Social semantics:
  - Winter break
Different Types of Multimedia Semantics

- Multimedia content semantics
- Multimedia situational semantics
- Multimedia retrieval semantics
- Multimedia social semantics

Technical perspective

Human perspective

- Common vacation
- cute
- Dolomiti
- person
Annotation modes

- Multimedia content semantics
- Multimedia situational semantics
- Multimedia retrieval semantics
- Multimedia social semantics

- Person
- Dolomiti
- Cute
- Common vacation

- Manual Class
- Feedback
- Capture (during creation)
- Extract from image data (ex. Viola-Jones)
Multimedia Semantics

Find me a good picture of Irene during our common vacation

- Manual Class
- Feedback
- Capture (during creation)
- Extract from (image) data
Core Questions for Multimedia

How to represent the ontology?
How to query the ontology?
How to construct the ontology?
How to populate the ontology?
**Ontologies**

“An Ontology is a formal specification of a shared conceptualization of a domain of interest” (Gruber, 93). An ontology is an engineering artifact, it is constituted by:

- a specific vocabulary used to describe a certain reality
- a set of explicit assumptions regarding the intended meaning of the vocabulary.

An Ontology is a formal specification of a shared conceptualization of a domain of interest.

Ontologies permit the formal representation of the knowledge domain:

- provide a formal and computable representation of a domain through the use of Semantic Web technologies, while standard classifiers provide only classification labels
- allow to represent concept properties and spatio-temporal relations between concepts.
Ontologies implement taxonomies. Taxonomy := Segmentation, classification and ordering of elements into a classification system according to their relationships between each other.

An ontology is composed by:

• **Classes**
• **Relationships between classes**
But also by:
- *Relationships between classes and data types*
And by: Individuals (or instances)
• Core questions:
  How to represent the ontology? *Ontology Representation Languages*
  OWL Lite, OWL DL, OWL Full, SWRL
  How to query the ontology? *Ontology Query Languages*
  SPARQL
  How to construct the ontology? *Ontology guidelines*
  How to populate the ontology? *Ontology-based*
  annotation
  content recognition
  content creation
  retrieval and feedback
  metadata management

How to perform reasoning?

SWRL permits to express rules explicitly through *if-then* expressions, and provides built-in mathematical, logical, string comparison and temporal operators, making it easier the definition and modification of rules and rule constraints

Reasoning services help knowledge engineers and users to:
  • Infer ontology class hierarchy, i.e. testing if a class is a subclass of another class (subsumption testing)
  • checking if a class can have any instance, based on the description of the class (consistency checking)
Ontology Representation Languages OWL

The Web Ontology Language (OWL) is a family of knowledge representation languages recommended by W3C for authoring ontologies. There are three variants:

- **OWL Lite** supports those user with classification hierarchy and simple constraint features. For example, while OWL supports cardinality constraints, OWL Lite only permits cardinality values of 0 or 1. OWL Lite is decidable.

- **OWL DL (OWL Description Logic)** includes all OWL language constructs with restrictions such as type separation (a class cannot also be an individual or a property; a property cannot also be an individual or class). OWL DL is decidable.

- **OWL Full** a class can be treated also as a collection of individuals and as an individual in its own right. OWL Full is not decidable. Typically serialized in RDF/XML
SPARQL

SPARQL is the query language for the Semantic Web RDF language, syntactically similar to SQL (W3C Recommendation)

Several implementations are already available:
  – Jena
  – Sesame
Some features are still missing (delayed for a next version):
  – sum, average, min, max functions …
  – aggregation of queries
  – …
SWRL (Semantic Web Rule Language)

SWRL (Semantic Web Rule Language) is intended to be the rule language of the Semantic Web. SWRL is based on OWL: all rules are expressed in terms of OWL concepts (classes, properties, individuals, literals...)

SWRL permits to express rules explicitly through *if-then* expressions, and provides built-in mathematical, logical, string comparison and temporal operators, making it easier the definition and modification of rules and rule constraints.

Reasoning services help knowledge engineers and users to build and use ontologies allowing to:

- Infer ontology class hierarchy, i.e. testing if a class is a subclass of another class (subsumption testing)
- Consistency checking, i.e. checking if a class can have any instance, based on the description of the class

Several reasoning tools available: Pellet, Kaon2, RacerPro, …
Ontologies for video annotation and retrieval

Ontologies can be regarded as an appropriate tool to bridge the semantic gap between the information that can be extracted from the raw video and its interpretation in a given context.

Traditionally ontologies for semantic video annotation and retrieval include only linguistic terms appropriate to distinguish categories of both concrete objects and events. Linguistic terms also express higher level abstract concepts of the domain.

Ontologies are designed to serve:
- Content analysis and annotation
- Semantic content-based retrieval
- Reasoning over content
- Personalized filtering of content
With ontologies classifiers are still used to instantiate the concepts in the ontology. Ontology-based reasoning (using concepts and feature values) then can be used to exploit contextual information, such as temporal constraints, to disambiguate situations in which visual feature values are not enough discriminative.

For example in soccer video domain a scored goal sequence may not be recognized using only the visual features (using ball tracking), but considering that shot on goal sequences are followed by cheering crowd and player close-up shots the highlight may be recognized using SWRL rules that model the event.
Solutions with Ontologies and classifiers

- **Snoek et al.** (TMM 2007) perform concept annotation of video clips with the MediaMill 101 concept lexicon. Machine learning is used to train classifiers that detect high-level concepts from low-level features. WordNet is used to derive high-level concepts relations to improve the performance.

- **Zha et al.** (ACM MIR 2007) An ontology was defined to provide some structure to the LSCOM-lite lexicon, expliciting pairwise correlations and hierarchical relationships between concepts. The ontology is used to refine concept detection of SVM classifiers.

- **Smeaton et al.** (TMM 2007) showed that inter-concept relationships are an important resource for semantic analysis of multimedia data and defined a methodology for the analysis of low-level features and semantic properties of three flat concepts lexicons (LSCOM, LSCOM-Lite, CDVP-206)
A few researchers have included visual knowledge in the ontologies. In this way ontologies also include instances that serve to verify conditions and constrains useful for classification:

- **Kompatsiaris, Staab, Strintzis et al.** (TCSVT 2005) defined a *Visual Descriptors ontology*, a *Multimedia Structure ontology* and a *Domain ontology* to perform video content annotation at semantic level. The Visual Descriptors ontology included concept instances represented with MPEG-7 visual descriptors.

- **Dasiopoulou et al.** (Springer 2008) proposed an ontology-based framework for enhancing segment-level annotations resulting from typical image analysis, through the exploitation of *visual context and spatial relations*.

- **Jain et al.** (ACM MM 2007) have employed a two-level ontology of artistic concepts that includes visual concepts such as color and brushwork in the first level, and artist name, painting style and art period for the high-level concepts of the second level. An inference framework has been used to annotate and disambiguate high-level concepts.
The Pictorially Enriched Ontology model includes concepts, concept instances, SWRL rules that describe more complex events, and a method to learn set of first-order logic rules that describe events defined in the ontology.
Implementation

The ontology includes media representations of perceptual facts like images or video clips (*perceptual concepts*) and link these elements to the *abstract concepts*. Perceptual concepts are *instances* of the corresponding abstract concepts. They provide a matching reference for the visual descriptors of the observed entities, and permit reasoning on instance values as well.

Three mechanisms are included in the ontology to support the effective annotation of entities with large intra-shot changes:

- **Visual instance clustering**: to cluster visual instances of similar patterns.
- **Cluster prototyping**: to select one or more visual representatives of each cluster, to reduce the cost of matching
- **Dynamic cluster updating**: to update prototypes whenever new knowledge is presented in ontology.

Finally, SWRL is used to perform reasoning over both concepts and concept instances, so as to disambiguate the results of the classification or derive new semantic annotations of parts of the clips.
Modeling the domain knowledge

The Pictorially Enriched Ontology model (in OWL) includes concepts and their visual descriptors. SWRL rules describe more complex events.
Core Questions for Multimedia:

- How to represent the ontology?
- How to query the ontology?
- How to construct the ontology?
- How to populate the ontology?

• Ontology Representation Languages

• Ontology Query Languages

• Ontology guidelines

• Ontology-based
  - annotation
  - content recognition
  - content creation
  - retrieval and feedback
  - metadata management
Ontology – In computer science

“People can’t **share knowledge** if they do not speak a **common language**”

[Gruber, 1993]

**Gruber 93:**

An Ontology is a

- formal specification → Executable, to discuss
- of a shared → group
- conceptualization → about concepts
- of a domain of interest → Set of information relative to a specific context
Ontology - Taxonomy

Taxonomy := Segmentation, classification and ordering of elements into a classification system according to their relationships between each other.
An ontology is composed by:
- **Classes**
- **Relationships between classes**
- **Relationships between Classes and DataTypes**
- **Individuals**
An ontology is composed by:
- Classes
- Relationships between classes
- Relationships between Classes and DataTypes (i.e. String, integer)
- Individuals
An ontology is composed by:
- Classes
- Relationships between classes
- Relationships between Classes and DataTypes
- Individuals
Ontologies for video semantic annotation and retrieval

• Visual features used for the creation and maintenance of the multimedia ontology

1. **Low-level perceptual features**, i.e. generic visual descriptors such as color histograms, edge maps, etc. that are usually related to generic concepts such as scene settings, shot type, classification of indoor/outdoor scene, etc.

2. **High-level semantic features**, i.e. high level descriptors such as face detection and recognition, superimposed text, etc. and specific descriptors that are usually related to the domain such as playfield detection and recognition in sports videos, anchorman detection in news and target tracking in video surveillance.
A multimedia ontology for the soccer domain

- The ontology is composed by a domain specific and a video structure ontology.

- The domain ontology includes concepts and relations that define the soccer domain.

- The video structure ontology describes the component elements of a video
A multimedia ontology for the soccer domain: the domain ontology

- Contains:
  1. all the concepts and relations that define the soccer domain (expressed in linguistic terms)
  2. the perceptual elements that model the visual patterns of the specializations of the concepts of the linguistic part (visual concepts)

- To account for the many perceptually different patterns in which they can manifest, visual concepts are clustered according to the similarity of their spatio-temporal patterns.

- For each cluster, a visual prototype is obtained as the representative element for that pattern.
A multimedia ontology for the soccer domain: the video structure ontology

- The visual prototype acts as a bridge between the domain ontology and the video structure ontology.

- A visual prototype is a structural element of the video and at the same time is linked to a linguistic concept of the domain ontology (a play action, a player).

- This link plays an important role for automatic annotation of new video sequences and for retrieval by content.
MPEG-7 and Ontology

• You learned something about the MPEG-7.

• How can we describe better its location in the multimedia retrieval field?
  – Former situation: no standard vocabulary
  – Current situation: MPEG-7; standard vocabulary, agreement between involved parties on which vocabulary to use and its exact meaning. The advantages of using MPEG-7 in multimedia domain is due to the fact that it has been designed to fully describe multimedia document structure, but at the same time it reflects the "structural" lack of semantic expressiveness of XML.
  – Future situation: standard vocabulary with pre-defined meaning
Meaning of Information:
(or: what it means to be a computer)

In the context of the page, the content appears to be discussing the meaning of information, possibly related to computer science or technology. The text is not fully legible due to the format of the page and the images. However, it appears to touch on themes of education, work, private, CV, and name, and includes some Chinese characters that suggest a focus on personal information or a resume. The content seems to be a mixture of English and Chinese, possibly discussing the significance of information in a global or multicultural context.
XML ≠ Meaning, XML = Structure

在愛戴者熱心奔走之下，華裔名指揮家林克昌根留台灣的可行性又提升了幾分。兩廳院主任李炎、國家音樂廳樂團副團長黃奕明日前親赴林克昌、石聖芳寓所拜會，並提出多場客席邀約。此外，台灣省立交響樂團團長陳澄清也茫茫「下訂」，邀請林克昌赴台中霧峰，從八月十日起訓練省交，為期長達一個月。
在台灣諸多公家樂團中，陳澄清是以實際行動表達對林克昌肯定的樂界人士之一，曾多次公開表示對林克昌指揮才華的欽佩，而且幾乎每個樂季都邀請林克昌客席演出。
此外，林克昌上個月赴俄羅斯與頂尖的「俄羅斯國家管絃樂團」灌錄了柴可夫斯基晚期三大交響曲以及「羅密歐與茱麗葉」、「斯拉夫進行曲」、「義大利隨想曲」，最後的DA T母帶也在前兩天寄回台灣。製作人楊忠衡與林克昌試聽之後，都對錄音效果一尤其音質表現感到相當滿意，楊忠衡估計呈現了十分林克昌指揮神韻。
俄羅斯國家管絃樂團首席畢魯尼日前也讚異林克昌的指揮藝術有三大特點：一是控制自如的彈性速度；二是強烈的動態對比；三是宛如呼吸歌唱的旋律處理。這些對錄音師而言都構成很大挑戰。俄國錄音師雖然採用多軌混音，但定位、場面都有可觀之處。
What is the Problem with MPEG-7?

How do you formulate a query to get all segments that show “Sky“?

First Shot:

XQL: //StillRegion[.//Keyword="Sky"]
What is the Problem with MPEG-7?

• Annotations are not interoperable:
  - Multiple ways to model semantically identical descriptions!
  - Complex queries needed to cover all alternatives!
How to overcome the drawbacks of MPEG-7?

• 1-to-1 translations from MPEG-7 to Ontology language (e.g. [Hunter, 2003a]) will not result in high quality ontologies!

• Replace MPEG-7 with a high quality multimedia ontology that fulfils the following requirements:
  – **Reusability**
    Design a core ontology for any multimedia related application
  – **MPEG-7-Compliance**
    Support most important description tools (decomposition, visual / audio descriptors, … )
  – **Extensibility**
    Enable inclusion of further
    • description tools (even those that are not part of MPEG-7!)
    • media types
  – **Modularity**
    Enable customization of multimedia ontology
Ontologies – Some Examples

• General purpose ontologies:

• Multimedia Ontologies
  – Acemedia harmonization effort:
    http://www.acemedia.org/aceMedia/reference/multimedia_ontology/

• Domain and application-specific ontologies:

• Semantic Desktop Ontologies
  – Semantics-Aware instant Messaging: SAM Ontology,
    http://www.uni-koblenz.de/FB4/Institutes/IFI/AGStaab/Research/sam
  – Haystack,

• Web Services Ontologies
  – Core ontology of services http://cos.ontoware.org
  – OWL-S, http://www.daml.org/services/owl-s/1.0/

• Ontologies in a wider sense
Ontology

RDF

Giuseppe Serra
http://www.micc.unifi.it/serra
serra@dsi.unifi.it
RDF Data Model

• **Resources**
  – A resource is a thing you talk about (can reference)
  – Resources have URI’s (Uniform Resource Identifier)
  – RDF definitions are itself Resources (linkage)

• **Properties**
  – slots, defines relationship to other resources or atomic values

• **Statements**
  – “Resource has Property with Value”
  – (Values can be resources or atomic XML data)
A simple Example

• **Statement**
  – “Irene is the creator of the resource http://www.w3.org/Home/Irene”

• **Structure**
  – Resource (subject) http://www.w3.org/Home/Irene
  – Property (predicate) http://www.schema.org/#Creator
  – Value (object) “Irene”

• **Directed graph**

http://www.w3.org/Home/Irene ➔ s:Creator ➔ Irene
Another Example

- To add properties to Creator, point through an intermediate Resource.

http://www.w3.org/Home/Lassila

s:Creator

Person://fi/654645635

Name

Irene

Email

irene@w3.org
Collection Containers

• Multiple occurrences of the same PropertyType doesn’t establish a relation between the values
  – The Millers own a boat, a bike, and a TV set
  – The Millers need (a car or a truck)
  – (Sarah and Bob) bought a new car

• RDF defines three special Resources:
  – Bag unordered values rdf:Bag
  – Sequence ordered values rdf:Seq
  – Alternative single value rdf:Alt
  • Core RDF does not enforce ‘set’ semantics amongst values
Example: Bag

- The students in course 6.001 are Amy, Tim, John, Mary, and Sue
Example: **Alternative**

- *The source code for X11 may be found at ftp.x.org, ftp.cs.purdue.edu, or ftp.eu.net*
Statements about Statements

- Making statements about *statements* requires a process for transforming them into Resources
  - **subject** the original referent
  - **predicate** the original property type
  - **object** the original value
  - **type** rdf:Statement
Example: Reification

• *Jonathan believes that*
  – *the creator of the resource* http://www.w3.org/Home/Irene *is Irene*
RDF Syntax I

- Datamodel does not enforce particular syntax
- Specification suggests many different syntaxes based on XML
- General form:

```xml
<rdf:RDF>
  <rdf:Description about="http://www.w3.org/Home/Irene">
    <s:Creator> Irene </s:Creator>
    <s:createdWith rdf:resource="http://www.w3c.org/amaya"/>
  </rdf:Description>
</rdf:RDF>
```
<rdf:RDF>
    <rdf:Description about="http://www.w3.org/Home/Irene">
        <s:Creator> Irene </s:Creator>
        <s:createdWith rdf:resource="http://www.w3c.org/amaya"/>
    </rdf:Description>
</rdf:RDF>
RDF Syntax II: Syntactic Varieties

Typing Information

<s:Homepage rdf:about="http://www.w3.org/Home/Lassila"
  s:Creator="Ora Lassila"/>

<s:Title>Ora's Home Page</s:Title>
<s:createdWith>
  <s:HTMLEditor rdf:about="http://www.w3c.org/amaya"/>
</s:createdWith>
</s:Homepage>

Subject (OID)

http://www.w3.org/Home/Lassila
Ora Lassila

In-Element Property

rdf:type
Property

s:Creator
s:createdWith
RDF Schema (RDFS)

• RDF just defines the datamodel
• Need for definition of vocabularies for the datamodel - an Ontology Language!
• RDF schemas are Web resources (and have URIs) and can be described using RDF
RDF-Schema: Example

\[ s = \text{rdfs:subClassOf} \]
\[ t = \text{rdf:type} \]
Rdfs:subclassOf

- `<rdfs:description about="Xyz:Minivan">`  
  `<rdfs:subclassOf about="xyz:Van"/>`  
  `</rdfs:description>`

- `<rdfs:description about="myvan">`  
  `<rdf:type about="xyz:MiniVan"/>`  
  `</rdfs:description>`

- **Predicate Logic Consequences:**

  - Forall X: type(X,MiniVan) -> type(X,Van).
  - Forall X: subclassOf(X,MiniVan) -> subclassOf(X,Van).
**Rdf:property**

- `<rdf:description about=„possesses“>`
  - `<rdf:type about=„....property“/>`
  - `<rdfs:domain about=„person“/>`
  - `<rdfs:range about=„vehicle“/>`
- `</rdf:description>`
- `<rdf:description about=„peter“>`
  - `<possesses>petersminivan</possesses>`
- `</rdf:description>`

**Predicate Logic Consequences:**

- Forall X,Y: possesses (X,Y) -> (type(X,person) & type(Y,vehicle)).
Links RDF

- RDF
  - RDF Primer:
    - http://www.w3.org/TR/rdf-primer/
  - RDF Concepts and Abstract Syntax:
    - http://www.w3.org/TR/rdf-concepts/
  - RDF Semantics Document
    - http://www.w3.org/TR/rdf-mt
  - RDF Vocabulary Description Language 1.0 RDF Schema:
    - http://www.w3.org/TR/rdf-schema/
  - RDF Viewer
    - IsaViz
      - http://www.w3.org/2001/11/IsaViz/
    - Protégé
      - http://protege.standford.edu