Ontology for Multimedia Applications

Hiranmay Ghosh
TCS Innovation Labs, Delhi
Contributors

Santanu Chaudhury
IIT Delhi

Anupama Mallik
IIT Delhi

Hiranmay Ghosh
TCS / IIT Delhi

... and many other students
Agenda

Part I

● Introduction
● Semantic Web and Ontology
● Multimedia Content Processing
● Ontology for Multimedia Data Interpretation

Part II

● Multimedia Web Ontology Language
● Application Examples
● Distributed Multimedia Applications
● Conclusion
Part I
Agenda

Part I
- Introduction
  - Semantic Web and Ontology
  - Semantic Multimedia Content Processing
  - Ontology for Multimedia Data Interpretation

Part II
- Multimedia Web Ontology Language
- Application Examples
- Distributed Multimedia Applications
- Conclusion
Multimedia for infotainment
Some statistics [2012]  

**Video**

- **14 million** – Number of Vimeo **users**.
- **200 petabytes** – Amount of **video played** on Vimeo during 2012.
- **150,648,303** – Number of unique visitors for video to Google Sites, the number one **video property** (September).
- **1 billion** – PSY’s Gangnam Style video became the first online video to reach 1 billion views (currently just over 1.1 billion) and it achieved it in just 5 months.
- **2.7 billion** – Number of views of videos **uploaded** to YouTube tagged Obama or Romney during the 2012 U.S. election cycle.
- **2.5 million** – Number of hours of **news-related video** that was uploaded to YouTube.
- **8 million** – The number of concurrent viewers of the live Baumgartner’s jump from the edge of space, the most ever.
- **4 billion** – Number of **hours of video** we watched on YouTube last year (September 2012).
- **60 million** – Number of **global viewers** monthly on Ustream.
- **16.8 million** – Number of **total viewers** in a 24 hour period for the most ever.
- **181.7 million** – Number of total unique viewers of online videos during **December**.

**Images**

- **7 petabytes** – How much photo content Facebook **added** every month.
- **300 million** – Number of new photos **added** every day to Facebook.
- **5 billion** – The total number of photos uploaded to Instagram since its start, reached in September 2012.
- **58** – Number of photos uploaded **every second** to Instagram.
- **1** – Apple iPhone 4S was the most popular **camera** on Flickr.
How do we deal effectively with the large volume of distributed multimedia data?

Organize
Retrieve
Navigate
Correlate
News aggregation

- TV Channels
- Newspapers
- Social Media
- Maps

- Aggregate
- Summarize
- Present
- Navigate

- Speech
- Video
- Overlay Text
- Image
- Text

November 19, 2013 Tutorial: Web Intelligence 2013
Digital Heritage

- Dance forms
- Music genres
- Instruments
- Myth
- Scripture
- Artistes
- Schools ...

- Videos
- Still images
- Document images
- Text

- Retrieve
- Navigate
Agenda

Part I
- Introduction
- **Semantic Web and Ontology**
  - Multimedia Content Processing
  - Ontology for Multimedia Data Interpretation

Part II
- Multimedia Web Ontology Language
- Application Examples
- Distributed Multimedia Applications
- Conclusion
The Semantic Web

- Semantic data modeling
  - Concepts represented through symbols
  - Relations between the concepts

- Common reference for interpretation of data from multiple sources

- Layers for
  - Syntactic compatibility (XML)
  - Semantic interoperability (RDF, OWL)
Ontology

- A formal representation of a domain

- An *artiste* is a *person*
- A *person* has name [string]
- A *dancer* is an *artiste*
- A *dancer* performs *dance*
- *DancerX* is a *dancer*
- *Bharatnatyam* is a *Dance*
- *DancerX* performs *Bharatnatyam*
- *DancerX* has name “Yamini Krishnamurthy”
Why use ontology?

- Template for information extraction
  \[<\text{dancer}> <\text{name}> <\text{dance-type}>\]

- Reasoning to find new facts (not explicitly stated)
  - \textit{DancerX} is a \textit{person}
  - \textit{DancerX} performs \textit{Dance}
  - At least one dancer performs \textit{Bharatnatyam}

- Separation of knowledge from program logic facilitates
  - \textit{Knowledge Engineering}
  - \textit{Reuse and maintenance}
Agenda

Part I

- Introduction
- Semantic Web and Ontology
  - Multimedia Content Processing
    - Ontology for Multimedia Data Interpretation

Part II

- Multimedia Web Ontology Language
- Application Examples
- Distributed Multimedia Applications
- Conclusion
Content, Concept & Context

- Content based retrieval (early 1990's)
  - Low level image features, e.g. Color & texture
- Concept based (late 1990's – still evolving)
  - Features conveying more semantics, e.g. SIFT
  - Machine Learning techniques
- Contextual reasoning
- Granularity of semantics
  - Scene recognition
  - Object recognition
    - Generic & Specific

A beach scene

- Water (Blue)
- Sky (Blue)
- Sand (Brown)
Current state of content understanding

- Significant progress in visual data understanding
  - Document images, Surveillance, Medical / Satellite imagery, Scene understanding, Action recognition, ...
- Audio & Speech
  - Good progress
- Domain specific solutions
  - Implicit domain knowledge
Semantic gap: still an unsolved problem

Semantic World

Semantic Gap

Media World

Bananas

Red Light

Bharatnatyam

STOP!

Joy and freedom
Agenda

Part I
- Introduction
- Semantic Web and Ontology
- Multimedia Content Processing
- **Ontology for Multimedia Data Interpretation**

Part II
- Multimedia Web Ontology Language
- Application Examples
- Distributed Multimedia Applications
- Conclusion
Multimedia Data Integration

- Different Media types
- Diversity in descriptors
- Difference in indexing schemes

Can we do semantic modeling of multimedia data?
Working with the annotations

- Multimedia data is often associated with annotation
  - Structured metadata, User tags, HTML `<ALT>` tag, surrounding text, ...
- We can use ontology to interpret them?
- A set of collaborating museums
  - Well-curated media archives
  - Controlled metadata associated with media artifacts
- OWL-based domain ontology for information integration

- Unfortunately, it does not work with any arbitrary media collection
Crowd-sourced data and knowledge  (2008 onwards)

- Semantics extracted
  - From Crowd-sourced tags
  - With Crowd-sourced knowledge (Wikipedia)
  - A new line of research

- But ...
  - Estimated 70% of social media contents are without tags
  - Automatic tagging
“Qualities” of concepts

- "Qualities": perceptible/measurable
  - Physical (color, size ...)
  - Relations (Spatial and temporal)
- Relation between and quality regions (qualia)
  - “Red” is opposite to “green”
  - “Red” is close to “brown”

Source: Gangemi (2002)
Multimedia Content Description Scheme

ISO Standard: MPEG-7: Early 2000's

- Flexible language to describe multimedia contents
  - Representations (tools) for common audio and visual features
    - Color, texture, shape, frequency spectrum, etc.
  - Scene description
    - Structural and semantic description
  - Extensible
    - Possible to define new descriptors
Description of still image

- **Still region SR1**: Creation information
  - Textual annotation

- **Spatial segment decomposition**: No overlap, gap

- **Still region SR2**: Textual annotation
  - Color Structure

- **Still region SR3**: Textual annotation
  - Color structure
  - Contour shape

- **Directional spatial segment relation**: left

- **Event EV1**: Label
  - Semantic time
  - Semantic place

- **Concept C1**: Label
  - Property
  - Property

- **Concept-semantically base relation**: hasPropertyOf

- **Object-event relation**: hasAgentOf

- **Object-event relation**: hasAccompanierOf

- **Comradeship**
- **Shake hands**
- **Alex**
- **Ana**

- **Agent object AO2**: Label
  - Person
Video segments

Video segment 1: Pass

Video segment 2: Kick and score

Moving region: Players
Moving region: Goal keeper
Moving region: Ball
Still region: Goal

Segment-Relationship Graph

Is composed of

Ball
Player 1
Player 2

Is close to
Right of
Moves toward

Same as

Video segment: Kick and score

Is composed of

Ball
Goal keeper
Player 2
Goal

In front of
Moves toward
Comments on MPEG-7

- **Accomplishes syntactic interoperability for multimedia**
- Describes multimedia document content
  - XML based schema
  - Lots of flexibility (same scene can be described in many different ways)
  - No semantics, no support for reasoning

- Quite a few MM Information system built with MPEG-7
  - Template matching (query by example paradigm)
Ontology for multimedia “concepts”

- Controlled vocabulary for MPEG-7 semantic description
  - Utility
  - Coverage
  - Feasibility
  - Observability

Source: Naphade (2006)
MPEG-7 Ontologies

- To provide semantic rigor to MPEG-7 descriptors
- Several research projects
  - Harmony
  - AceMedia
  - DS-MIRF
  - COMM
  - Boemie
  - ...
- Converts MPEG-7 constructs to RDF / OWL constructs
- Different coverage to MPEG-7 parts
MPEG-7 Ontology: Class hierarchies

Top level content entities

Segment classes
MPEG-7 Ontology: Media Properties

- Visual Descriptor
- subclass of
- Color
- Range
- MPEG-7 Color Tools
- subclass of
- Color Layout
- Scalable Color
- Dominant Color
- Still region
- Video segment
- Domain
MPEG-7 Ontologies ... contd.

- Creates semantic description of multimedia contents in collections
  - Excludes semantic descriptors
    - Integrates with domain ontology
      - Usually with a core ontology
    - Examples: Harmony, AceMedia, COMM
  - Includes semantic descriptors
    - Results in independent semantic descriptions of repositories
    - Needs common understanding of domain
    - Example: DS-MIRF
Interoperability models

MPEG-7

MPEG-7 Ontology
- Harmony
- AceMedia
- COMM

Source: Dasiopoulou (2010)
Architecture for semantic integration

- Concept Interpretation and Fusion
  - Domain knowledge
  - Multimedia Ontology
    - Media Models
- Media Processing
  - Image
  - Video
  - Video OCR
  - Text
  - Speech
Comments on MPEG-7 ontologies

- Content model for documents / collections
  - Can correlate diverse media forms

- Specific to multimedia instances
  - Not a generic collection independent ontology

- Media model and domain model form separate layers
  - Media interpretation does not benefit from domain knowledge
Pictorially enhanced ontology

Univ Fierenze: Mid-Late 2000's

- Visual templates (examples) are associated with media events (concepts)
  - Each template represents a distinct modality of manifestation
- New instances are classified based on feature similarity with prototypes
  - Automatic event detection and annotation
- Domain ontology relates such events
Pictorially enhanced ontology ... contd

End of Part I