A framework for building Multimedia Ontologies from Web Information Sources

ANGELO CHIANESE, VINCENZO MOSCATO, FABIO PERSIA, ANTONIO PICARIELLO, CARLO SANSONE

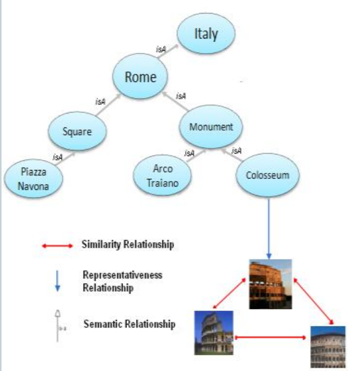
DIS - UNIVERSITY OF NAPLES
VIA CLUADIO 21, 80125, NAPOLI, ITALY
{angchian,vmoscato,fabio.persia,picus,carlosan}@unina.it

Introduction

- The Web 2.0 has changed the relation between users and Internet…
  - a lot of repositories (e.g. Flickr, YouTube, etc.) containing both multimedia data and the related annotations or metadata are publicly available on the web
- Usually, information are described by means of "flat" metadata or sometimes using small annotations in natural languages
  - such a kind of structures are substantially inadequate to support complete retrieval by content of multimedia documents

Research Issues

- The main idea beyond this work is that such a kind of information can be efficiently used for an automatic extraction of *multimedia knowledge*, particularly suitable for a variety of applications…
  - *How to represent, organize and manage multimedia data and the related semantics by means of a formal framework?*
  - *Is the use of* ***multimedia ontologies*** *the solution?*



Addressed Research Questions

- *What a multimedia ontology is: is it a taxonomy, or a semantic network of metadata (tags, annotations)?*
- *Does a multimedia ontology support concrete data: what is the role of rough data – image, video, audio data– if any?*
- *What a multimedia semantics is: how to define and capture the semantics of multimedia data?*
- *How to build extensional ontologies: once defined a suitable formal framework, can we automatically build the defined multimedia ontologies?*

Our proposal

- The original contribution of this work is:
  - to propose a novel multimedia ontology framework, (MOWIS) especially in the image domain
  - to propose a technique for building ontologies, that operates on large corpora of human annotated repositories, considering both low level image processing strategies and keywords and annotations produced by humans
- The system supporting the described framework has been realized within the PRIN 2007-2009 project *Cooperare* and presented in previous work

Related Work (1/2)

- Multimedia Ontologies represent a way to formally specify the knowledge related to a specific domain by means of multimedia documents
  - …they are able to model a domain knowledge exploiting low-level features, structure, semantic concepts of multimedia data and the related relationships
  - Usually low-level features are machine-oriented and can be automatically extracted while semantic concepts are manually provided and are meaningful information only in specific domains
- Multimedia ontologies should allow the mapping between low-level and high-level information of multimedia data or their parts, thus supporting a more effective retrieval
- The problem of the semantic annotation by means of a multimedia ontology was largely investigated in the literature, in the opposite the automatic building of multimedia ontologies still remains an open issue and a challenging task

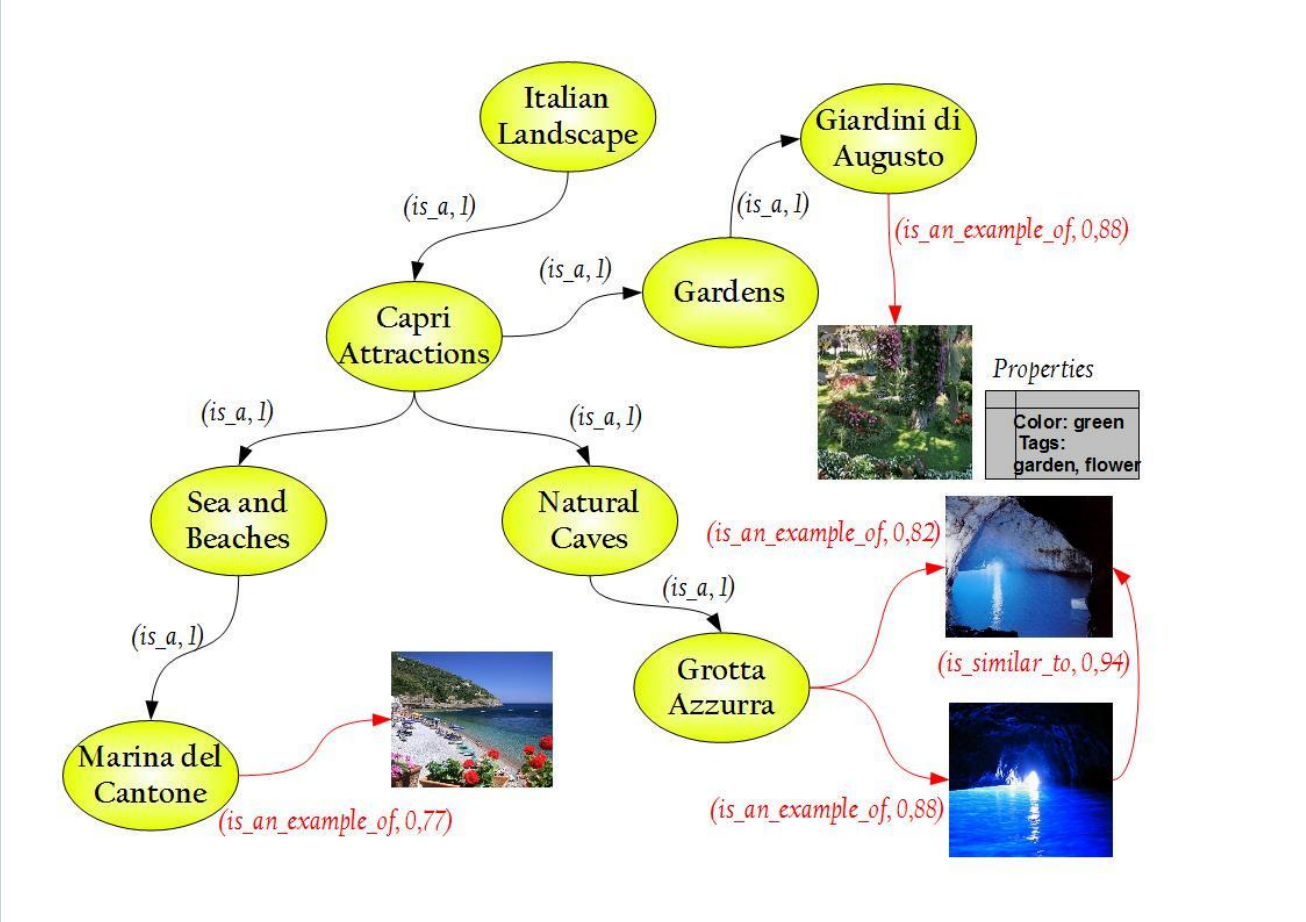
Related Work (2/2)

- The process for building multimedia ontology is structured into three steps:
  - selection of the concepts to be included in the ontology
  - definition of properties and relationships for the concepts
  - population and maintenance of the ontology
- *Concept-driven* vs *data-driven* approaches
- In the literature:
  - interesting techniques for extracting semantic concepts by clustering multimedia on the base of their visual features
  - evolutionary and incremental population of a multimedia ontology
  - semantic integration of metadata

An Image Ontology Model

- An Image Ontology can be modeled by a directed and labeled graph *(V, E,ρ)*, where:
  - *V={Vl ∪ Vh}* is a finite set of nodes formed by *low-level nodes Vl* (i.e. instances of images or image sub-regions, having specific properties such as content ore more enhanced features and general or domain-specific metadata) and high-level nodes *Vh* (i. e.g. concepts or instances of general, domain-specific or image content concepts)
- *E* is a subset of *(V x V)*
- *ρ* is a function that associates to each couple of nodes a label indicating the kind of relationship between the related classes *ρs* and its reliability degree *ρ*r
  - *Similarity relationship*, *Membership relationship*, *Representative relationship*, *Semantic relationship*

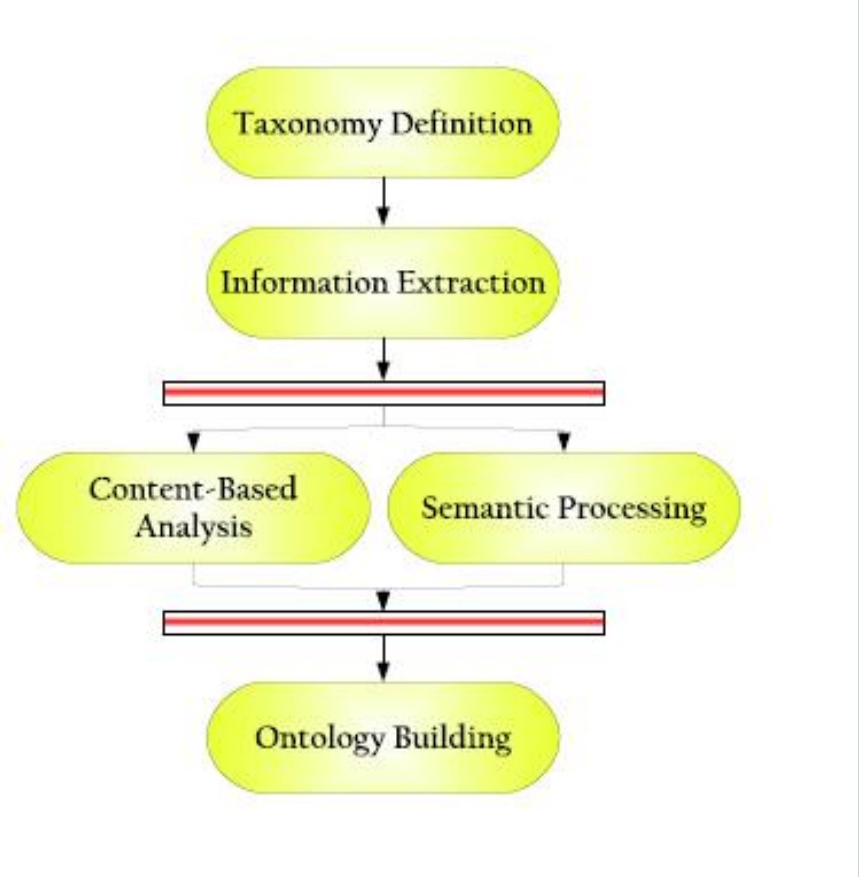
A graph representing the extensional part of an Image Ontology



The building process (1/3)

- The process useful to automatically build the graph representing our image ontology is based on a data-driven unsupervisioned approach and is composed by the following steps
  - *Definition of an initial taxonomy* containing a relevant concepts' instances hierarchy of the considered domain, that is represented by a subset of high level nodes
    - the definition is performed by experts in the domain of interest (domain-oriented approach)
  - *Information Extraction* (images and the related textual descriptions) from publicly available image repositories
  - *Content Based Analysis* of images and *Semantic Processing* of texts
  - *Ontology Bulding*

The building process (2/3)



The building process (3/3)

- The images are analyzed in order to obtain a low-level description using classical Computer Vision techniques
  - in particular the *Animate Vision* approach…
- Images derived from the same concept are then clustered (using the *EM algorithm*) to obtain their representative
- The textual part is at the same time processed in order to discover textual labels that better reflect image semantics using NLP techniques and topic detection algorithms
  - *Meta-Noise filtering*, *Named Entity Filtering*, *Linguistic Normalization and Filtering*, *Word Sense Disambiguation*, *Topic Detection*
- Onotolgy is then updated and more relevant topics (winner topics) derived from an image clutser are ‘’promoted” to be concepts in the ontology using the *Okapi BM25 ranking function*

Example of text processing

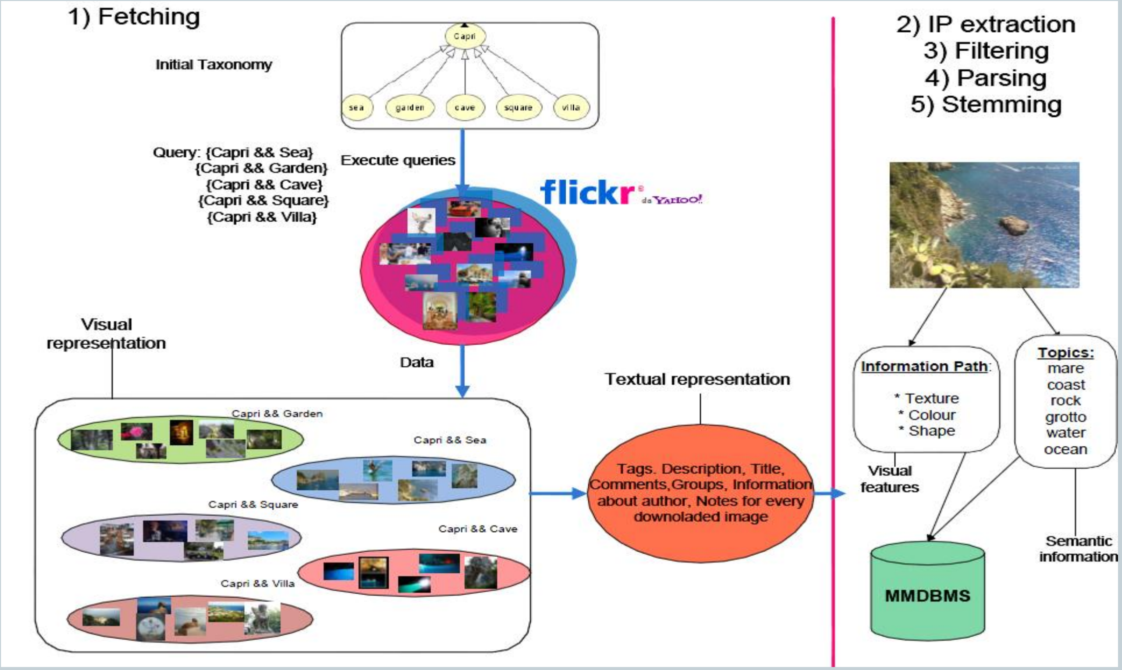
image	title	descr.	tags
	“Kruger Elephant”	“Bull elephant in Kruger National Park ”	{ Kruger National Park, Kruger, Elephant, bull, wildlife, Africa , South Africa, Portrait, landscape, BFGreatestHits }
	“Grand Street: Texting”	“This was actually taken at the corner Bowery and Delancy... but it was during my Grand Street excursion. ”	{ street, bicycle, woman, red shoes, cars, texting, morizariza, nyc, grand street, interestingness, New York, New York City, big apple, 100v+10f, blogged, modomatic, people, public, urbanphotooftheweek, WNYC, mo, mohammad, city, FlickAwardR, SuperbMasterpiece }

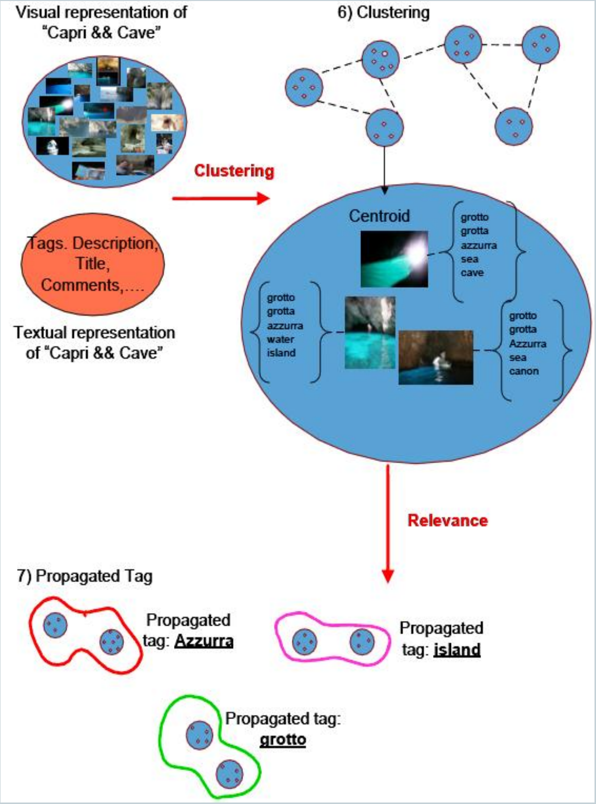
Table 1. Flickr images and their related textual information

Text Information	Meta-Noise Filter	Named Entity Filter	Linguistic Normalization	Linguistic Filter	Token.	WSD	Topic
‘Kruger Elephant”, “Bull elephant in Kruger National Park ”, { Kruger National Park, Kruger, Elephant, bull, wildlife, Africa, South Africa, Portrait, landscape, BF-Greatest Hits}	‘Kruger Elephant”, “Bull elephant in Kruger National Park ”, { Kruger National Park, Kruger, Elephant, bull, wildlife, Africa, South Africa, Portrait, landscape}	‘*location* Elephant”, “Bull elephant in *location* National Park ”, { *location* National Park, *location*, Elephant, bull, wildlife, Africa , South Africa, Portrait, landscape}	‘location elephant”, “bull elephant in location national park ”, { location national park, location, elephant, bull, wildlife, africa, south africa, portrait, landscape}	‘location elephant”, “bull elephant in location park ”, { location park, location, elephant, bull, africa , africa, portrait, landscape}	(⟨ location,4 ⟩, ⟨elephant, 2 ⟩ , ⟨ bull, 2 ⟩, ⟨ park, 2 ⟩ , ⟨ africa, 2 ⟩, ⟨ portrait, 1 ⟩, ⟨ landscape, 1 ⟩)	(⟨ location,4, syn_1 ⟩, ⟨ elephant,2, syn_2 ⟩ , ⟨ bull, 2, syn_3 ⟩, ⟨ park, 2, syn_4 ⟩ , ⟨ africa, 2, syn_5 ⟩, ⟨ portrait, 1, syn_6 ⟩, ⟨ landscape, 1, syn_7 ⟩)	(⟨ location,4, syn_1, 0.5 ⟩, ⟨ elephant,2, syn_2, 0.9 ⟩ , ⟨ bull, 2, syn_3, 0.1 ⟩, ⟨ park, 2, syn_4 ,0.6⟩ , ⟨ africa, 2, syn_5, 0.7 ⟩, ⟨ portrait, 1, syn_6, 0.4 ⟩, ⟨ landscape, 1, syn_7, 0.5 ⟩)

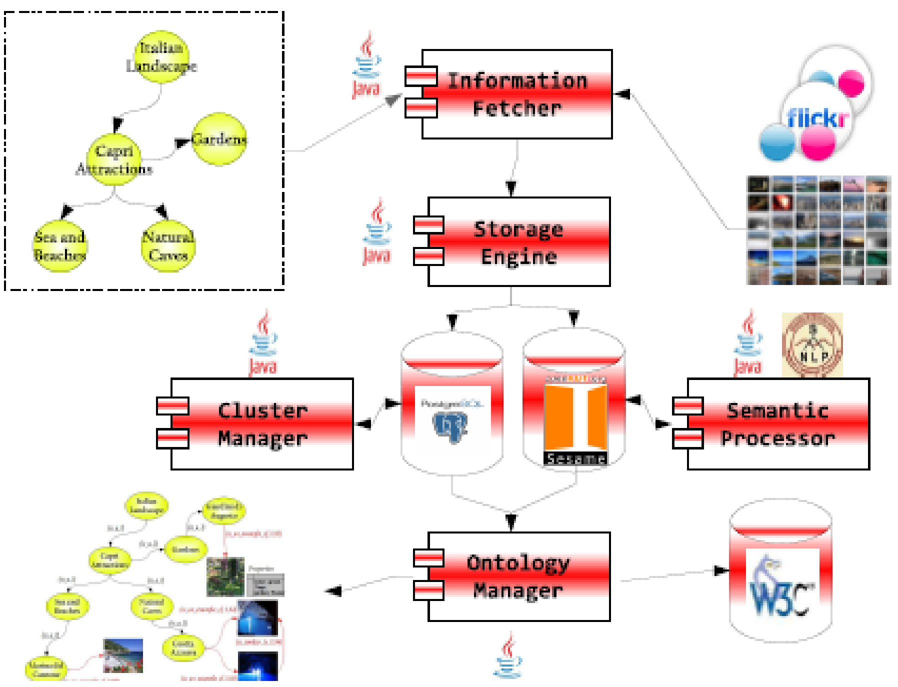
Table 2. Results of the semantic processing for the image in Table 1

Example of building process





The System Architecture



A Case Study and Preliminary Experimental Results (1/4)

- Evaluating the ''quality'' of an ontology is an important issue
- For this aim, we have built an ontology related to Capri, a wonderful Italian island of the Sorrentine Peninsula, on the south side of the Gulf of Naples
- A set of experts of natural and cultural attractions of Capri provided as initial taxonomy a graph containing the most relevant concepts in terms of high level nodes for the considered domain
- The FLICKR repository has been queried using as search keywords the logical AND among concepts reported in the leaf nodes of the taxonomy and the one corresponding to the root node

A Case Study and Preliminary Experimental Results (2/4)

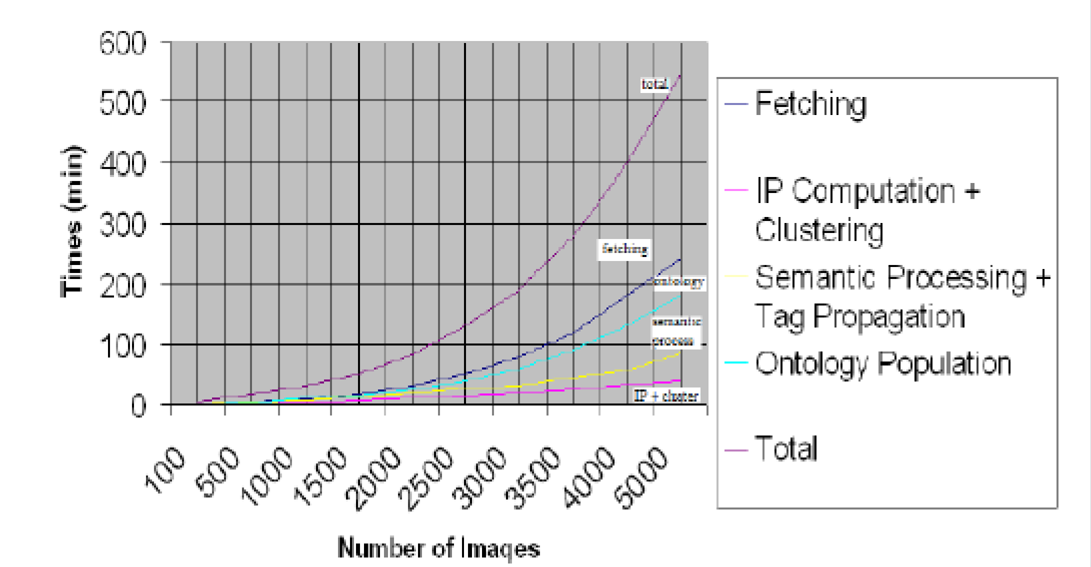
- Each retrieved image with the related annotations undergoes the described content-based analysis and semantic processing to determine the low-level description and the relevant labels to propagate in the ontology
- Our efforts have been then devoted to produce experimental results in order to evaluate the effectiveness of the produced ontologies with respect to some generated by human domain experts w.r.t. different criteria: Class Match Measure (CMM), DEnsity measure (DEM), Semantic Similarity Measure (SMM), BEtweenness Measure (BEM)
  - The Class Match Measure is meant to evaluate the “coverage” of an ontology for the given search terms
  - The DEnsity Measure is a metric that tries to measure the “representational density” or “informative content” of classes and consequently the level of knowledge detail
  - The Semantic Similarity Measure calculates how close the classes that matches the search terms are in an ontology
  - The BEtweenness Measure calculates the number of the shortest paths that pass through each couple of matched high-level nodes (betweenness) in the ontology

A Case Study and Preliminary Experimental Results (3/4)

- We ask five persons to describe in an exhaustive way and by means of an image ontology (concepts and photos selected from FLICKR) the main natural and cultural attractions of Capri, classifying them on the base of the related kind (sea and beaches, natural caves and gardens, squares and ancient villas)
- Then, we compared such ontologies with that one produced by our system using the knowledge (images, tags, description and titles) associated to the same photos (about 1000) from FLICKR.

A Case Study and Preliminary Experimental Results (4/4)

Ontology	CMM	DEM	SSM	BEM	Avg Score
MOWIS	1	0,89	0,64	0,578	0,777
User A	0,68	0,89	1	1	0,892
User B	0,76	1	0,87	0,59	0,805
User C	0,8	0,89	0,846	0,578	0,78
User D	0,7	0,89	0,87	0,52	0,745
User E	0,63	0,94	1	0,315	0,72



Conclusions and Future Work

- We addressed the problem of building a multimedia ontology in an automatic way using annotated image repositories
- Future work will be devoted to:
  - enlarge our experimentation to more significant case studies discussing the ontology maintenance problem and to make compatible output of the proposed framework with the latest languages for describing multimedia ontology