A Framework for Building Multimedia Ontologies from Web Information Sources*

Angelo Chianese, Vincenzo Moscato, Fabio Persia,
Antonio Picariello, and Carlo Sansone

DIS - University of Naples, via Cluadio 21, 80125, Napoli, Italy
{angchian,vmoscato,fabio.persia,picus,carlosan}@unina.it

Abstract. The definition of ontologies within the multimedia domain still remains a challenging task, due to the complexity of multimedia data and the related knowledge. In this paper, we present a novel framework (MOWIS) aiming at realizing a system for building Multimedia Ontologies from Web Information Sources that has been realized within the PRIN 2007-2009 project Cooperare and presented in previous works. In particular, we propose: i) a multimedia ontology model that combines both low level descriptors and high level semantic concepts; ii) automatic construction of ontologies using the FLICKR web services that provide images, tags, keywords and sometimes useful annotation describing both the image content and personal interesting information. Eventually, we describe an example of automatic ontology generation in a specific domain and present some preliminary experimental results.

1 Introduction

The Web 2.0 has changed the relation between users and Internet and nowadays, a lot of repositories containing both multimedia and the related annotations or metadata are publicly available on the web. 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, such as multimedia information indexing and retrieval, multimedia content visualization and browsing, learning, reasoning and so on. Indeed, information in the most diffused multimedia databases on the Web (e.g. Flickr, YouTube, etc...) is described by means of "flat" metadata, the most of the times using a predefined set of 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.

On the other side, it is well known in the literature that despite the tons of papers produced about multimedia databases and knowledge representation, there is not yet an accepted solution to the problem of *how to represent, organize and manage multimedia data and the related semantics by means of a formal framework*. It is the authors' opinion that there is still a great work to do with respect to the *intensional aspects* of a *multimedia ontology*: (i) *What a multimedia ontology is*: is it a taxonomy, or a semantic network of metadata (tags, annotations)? (ii) *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?

* Extended Abstract

Nicola Ferro and Letizia Tanca (Eds.): SEBD 2012, Edizioni Libreria Progetto, Padova, Italia
ISBN: 978-88-96477-23-6,

How to build extensional ontologies: once defined a suitable formal framework, can we automatically build the defined multimedia ontologies?

Throughout the rest of paper, we will try to give an answer to all the previous cited aspects; in particular the original contribution of this work is: (i) to propose a novel multimedia ontology framework, (MOWIS) especially in the image domain; (ii) to propose a technique for building ontologies, that operates on large corpora of human annotated repositories, namely the *FLICKR* database and the related *folksonomy* [8], considering both low level image processing strategies and keywords and annotations produced by humans when they store the produced data.

In the most common vision, *Multimedia Ontologies* represent a way to formally specify the knowledge related to a specific domain by means of the use of multimedia documents, such as images, videos, audio and texts. In particular, they are able to model a domain knowledge exploiting low-level features, structure, semantic concepts of multimedia data and the relationships (of different kinds) among them. Usually low-level features are machine-oriented and can be automatically extracted (as for *MPEG7* descriptors), 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 [7]. Respect to the problem of the *semantic annotation* by means of a multimedia ontology that was largely investigated in the literature [13], the automatic building of multimedia ontologies still remains an open issue and a challenging task. Generally, the process for building multimedia ontology is structured into three steps: (i) selection of the concepts to be included in the ontology, (ii) definition of properties and relationships for the concepts, (iii) population and maintenance of the ontology. To this aim, the main approaches for ontology building in the literature are those *concept-driven* and *data-driven*. In the first case the ontology is built by the knowledge related to a specific domain, while in the second case it is directly obtained from multimedia information, but a preliminary domain knowledge is used to select suitable data.

In the following we briefly describe the most diffused techniques in the literature for managing multimedia ontologies from which our work takes its roots. In [2] a first interesting technique for extracting semantic concepts by clustering images on the base of their visual and text features from annotated repositories is illustrated. While, a first proposal of how to combine visual and semantic descriptors to support annotation of multimedia contents in a specific domain is described in [4]. More recently, Bertini et al. presented in [3] the framework *MOM* (*Multimedia Ontology Manager*) based on the concept of *enriched ontology* for automatic video annotation and retrieval. Videos are grouped into clusters on the base of visual features and linked to specific semantic concepts (those corresponding to the objects that are representatives of the clusters). Moreover, facilities for expressing multimedia complex queries on the ontology are provided. In the opposite, *BOEMIE* [11] uses a supervisioned approach based on the extraction of semantic concepts from multimedia data to obtain in a evolutionary and incremental manner a multimedia ontology. Finally, *OntoMedia* [12] is a particular multimedia ontology framework that aims at managing very large collections of information by using techniques of *semantic integration* of metadata.

Nicola Ferro and Letizia Tanca (Eds.): SEBD 2012, Edizioni Libreria Progetto, Padova, Italia
ISBN: 978-88-96477-23-6,

In this paper, we first provide an algorithm for creating image ontology in a specific domain gathering high and low level multimedia information. We then give an example of automatic construction of image ontology and a discussion of the encountered problems and the provided solutions. Finally, we conclude showing preliminary experimental results. We want to remark that the system supporting the described framework has been realized within the PRIN 2007-2009 project *Cooperare* and presented in previous works [9, 10].

2 Building an Image Ontology

2.1 An Image Ontology Model

Stressing its conceptual nature, an ontology may be considered as a *theory* used to represent relevant notions about domain modeling in terms of concepts, relationships and constraints on them. Let us consider the image domain: as usual in a given media, we detect symbols, objects and concepts; in a certain image we have regions of pixels (symbol) related to portions of multimedia data; these regions are instances (object) of a certain concept. In other words, we can detect concepts but we are not able to disambiguate among the instances without some specific knowledge. A simplified version of the described vision process will consider only two main levels: *Low* and *High*. In fact, the knowledge associated to an image can be easily described at two different levels of analysis: (i) descriptions of raw images or of their parts(regions); (ii) general or domain-specific image content concepts that can be derivable or less from low-level.

Following such general considerations, an *Image Ontology* can be modeled by a directed and labeled graph $(\mathcal{V}, \mathcal{E}, \rho)$, where: (i) $\mathcal{V} = \{\mathcal{V}_l \cup \mathcal{V}_h\}$ is a finite set of nodes formed by *low-level nodes* $\mathcal{V}_l$ (i.e. instances of images or image sub-regions, having specific properties such as content - e.g. texture, shape, color, objects, etc... or more enhanced *features* and *general* or *domain-specific* metadata - e.g. author, title, description, tags, etc...) and *high-level nodes* $\mathcal{V}_h$ (i. e. instances of *general*, *domain-specific* or *image content* concepts); (ii) $\mathcal{E}$ is a subset of $(\mathcal{V} \times \mathcal{V})$; (iii) ρ is a function that associates to each couple of nodes a label indicating the kind of *relationship* between the related classes ρ_s, and its reliability degree $\rho_r \in [0, 1]$: $\rho : \mathcal{E} \rightarrow \langle \rho_s, \rho_r \rangle$.

Note that the effective use of this theory depends on the different kind of relationships that we can provide to the model, as described in the following. First, we provide a *similarity relationship*, that associates a couple of low-level nodes (images or sub-images) to a similarity degree, thus modeling the classical image matching algorithms based on low-level features (e.g. color, texture, shape, etc...) and/or semantic distances (e.g. taxonomy-based, Wu-Palmer, etc...) among metadata; a *membership relationship* verify if a given sub-image belongs to a given image; a *representativeness relationship* permits to associate those content features that better represent a given concept – using, for example, proper clustering or other classification algorithms that are able to determine what is the probability that an image is a valid representative of a concept; eventually, a *semantic relationship* may be applied to two high-level nodes (hypernym/hyponim, holonym/meronym, synonym, retrievable relations on lexical databases).

An example of a graph representing the extensional part of an image ontology related to some Italian landscapes is reported in Figure 1.

Nicola Ferro and Letizia Tanca (Eds.): SEBD 2012, Edizioni Libreria Progetto, Padova, Italia
ISBN: 978-88-96477-23-6,

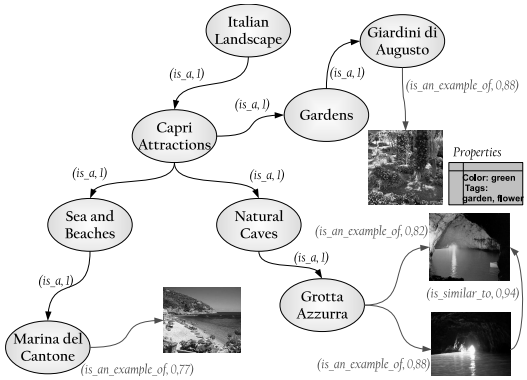

Fig. 1. An example of Image Ontology

2.2 The Building Process

As well noted in the literature, the critical part of an ontological framework is the effective construction of ontologies in a certain domain. Here, we describe a process useful to automatically build the graph representing our image ontology, by means of data-driven unsupervisioned approach. The proposed process, which is detailed in our previous works [9, 10], starts with the 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).

After the taxonomy definition, we first extract images and the related textual description from publicly available image repositories, such as FLICKR. The images are then analyzed in order to obtain a low-level description in terms of content features, using classical Computer Vision techniques; 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 on the textual annotations coming from the considered image repositories. In the image analysis task, we use a *salient points* technique - based on the *Animate Vision* paradigm [5] - that exploits color, texture and shape information associated with those regions of the image that are relevant to human attention, in order to obtain a compact characterization (*Information Path*) that could be also used to evaluate the similarity between images, and for indexing issues. Eventually, apposite super-visioned classification algorithms are exploited to determine content features [5].

We then analyze the textual information [9] that is usually associated to an image. To these purposes, information coming from *tags* are analyzed in combination with *titles* and *descriptions* by suitable NLP techniques that overcome the linguistic and semantic heterogeneity of such information, in order to extract a set of *relevant keywords* which more effectively represent image content. In particular, the semantic processing can be decomposed into a set of sequential tasks: (i) the *Meta-Noise Filtering* - aims at rejecting inaccurate or irrelevant words in the input text; (ii) the *Named Entity Filtering* - is used to find particular words (*named entities*) in the input text and associate to

Nicola Ferro and Letizia Tanca (Eds.): SEBD 2012, Edizioni Libreria Progetto, Padova, Italia
ISBN: 978-88-96477-23-6,

them the related semantic classes (person, organization, location, date, etc...); (iii) the *Linguistic Normalization* - performs the classical *stemming*; (iv) the *Linguistic Filtering* executes a *part of speech tagging* on the input text and selects only the words of *noun type*; (v) the *Word Sense Disambiguation (WSD)* - tries to automatically determine for each word the most suitable meaning by linking each word with the related WordNet synset; (vi) the *Topic Extraction* - aims at extracting a set of relevant keywords or *labels* with a *confidence* value considering both the *semantic similarity* among words, and the related *frequency* and *co-occurrence* of the words in the text.

Finally, the obtained knowledge in terms of images, low-level characteristics and labels is then merged into a graph that represents our image ontology. The proposed strategy is discussed in the following (see [10] for more details). Initially, all the images and the regions - whose relevant labels are associated to high-level nodes with a high grade of confidence - and that correspond to the leaves of the domain taxonomy - will be represented by apposite low-level nodes in the graph. In addition, couples of image nodes, whose similarity (computed by the *Information Path Matching* algorithm [5] and Wu/Palmer metric) is greater than a given threshold, will be connected by an edge having as reliability degree the related similarity measure. All the images belonging to the same concept are then clustered into different groups, which contain images that are more similar among themselves. We used as clustering procedure the *BEM* algorithm [5], that is recursively invoked to dynamically determine more fitting clusters without knowing a-priori the number of clusters themselves . Then we selected for each cluster the representative image as the closest one to all the other images of the cluster, and a suitable *representation probability* is associated to each representative image on the base of minimum and average distances. The process is iterated for each taxonomy leaf concept and the ontology is incrementally built: images belonging to different topics could be linked on the base of their similarity values allowing to *merge* the multimedia knowledge in a unique graph. Eventually, by a *Learning Tag Relevance algorithm* [6] (*Okapi BM25* ranking function), the topics that are more relevant – w.r.t. the content of images belonging to the same cluster (*winner topics*) – are *promoted* to be high-level nodes of the image ontology. The winner topics, whose relevance is greater than a threshold, are finally inserted as high-level nodes in the ontology and *linked*, from one hand to the image node that corresponds to the cluster centroid and, from the other one, to those nodes whose semantic distance (i.e. Wu/Palmer) is the minimum with respect to the current topic. If it is possible, the new ontology edge is labeled with the type of semantic relationship. In the case of topics that cannot be retrieved in WordNet, they are linked to the deepest high-level nodes of ontology from which they descend.

3 The System Architecture

The system architecture that supports the ontology building process is shown in figure 2. The user generates by a GUI an *OWL* file coding the initial taxonomy containing relevant concepts of the considered domain. Such a file is then the input of the *Information Fetcher* module that downloads images and the related annotations from the *Multimedia Repository*, using as *search keywords* the concepts related to the leaf nodes of the taxonomy and some filters on users.

Nicola Ferro and Letizia Tanca (Eds.): SEBD 2012, Edizioni Libreria Progetto, Padova, Italia
ISBN: 978-88-96477-23-6,

A *Storage Engine* module receives such information and stores image annotations (title, author, tags, labels, etc...) in a dedicated *RDF Database* and it stores both raw data and low-level characteristics in a *Image Database*. Finally, the *Semantic Processor* and *Cluster Manager* analyze high and low level information in order to generate/update, by means of the *Ontology Manager* and in according to the described process, a graph which represents the final multimedia ontology (stored in a OWL format in a dedicated repository). As for implementation details, we notice that: (i) the initial taxonomy is generated by a JAVA desktop application that uses *Protege'* API; (ii) FLICKR has been chosen as the multimedia repository; (iii) the Information Fetching module has been implemented as a JAVA application that exploits the FLICKR API; (iv) the RDF and Image Database have been realized by *Sesame* and *PostegreSQL* DBMS, respectively; (v) the Indexing packages have been implemented by apposite JAVA packages that exploit Stanford NLP and Animate Vision libraries.

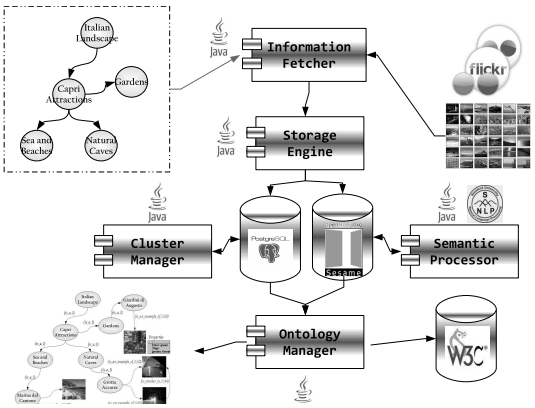

Fig. 2. System Architecture

4 A Case Study and Preliminary Experimental Results

Evaluating the "quality" of an ontology is an important issue both for ontology delevepers and for final users: this task allows to compare different ontologies describing the same domain in order to choose the more suitable one for a given application. 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 and exploiting some filters on user *id*s, in order to retrieve images really belonging to the domain.

Nicola Ferro and Letizia Tanca (Eds.): SEBD 2012, Edizioni Libreria Progetto, Padova, Italia
ISBN: 978-88-96477-23-6,

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*) [1]. To compute the different metrics, we ask five persons [1] 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).

Table 1. Values of the quality metrics for the *Capri* ontologies

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

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. For what the CMM metric computation concerns, we asked other 5 different user to provide a set of 50 common search terms for interesting Capri attractions (e.g. faraglioni, piazzetta, Jovis, marina grande, augusto, cave, azzurra, beach, anacapri) among FLICKR tags. Table 1 summarizes the experimental results for the different metrics obtained comparing the Capri ontology automatically generated by our system with respect to ontologies produced by the five human domain experts. The most expert users are indicated with A, B and C, while D and E represent the least experts. As we can see from the table, our ontology has a quality index very close to that of an ontology generated by humans quite experts on the considered domain. Finally, we measured the times (download, IP computation and clustering, semantic processing and tag propagation, ontology population) for building an image ontology depending on the number of fetched images [2]. We observed that The *Capri* Ontology, being composed by 1000 images, requires less than 10 minutes for its complete building (about two minutes if we do not consider the download from FLICKR), thus our apporach ensures a quite good scalability.

[1] In particular, we choose three persons more expert and other two less expert on the Capri attractions

[2] we use a Linux Ubuntu platform running on a 8GB RAM single CPU

Nicola Ferro and Letizia Tanca (Eds.): SEBD 2012, Edizioni Libreria Progetto, Padova, Italia
ISBN: 978-88-96477-23-6,

5 Conclusions

In this paper, first 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 (e.g. *M-OWL*).

6 Acknowledgments

The prototype realization has been carried out partially under the financial support of the FARO Programme in the framework of the *LINCE* Project: un sistema di Localizzazione/georeferenziazione degli INCidEnti stradali a basso costo.

References

1. Alani, H., Brewster , C.: Metrics for Ranking Ontologies. 4th Int. EON Workshop in 15th Int. World Wide Web Conf. (2006)
2. Benitez, A.B., Chang, S.F: Perceptual Knowledge Construction From Annotated Image Collections, International Conference on Multimedia & Expo. pp. 189-192 (2002)
3. Bertini, M., et al.: MOM: multimedia ontology manager. A framework for automatic annotation and semantic retrieval of video sequences, ACM International Conference on Multimedia, ACM Multimedia, pp. 787-788 (2006)
4. Bloehdorn, S., et al.: Knowledge representation for semantic multimedia content analysis and reasoning, Workshop on the Integration of Knowledge, Semantics and Digital Media technology (2004)
5. Boccignone, G., Chianese, A., Moscato, V., Picariello, A.: Context-sensitive Queries for Image Retrieval in Digital Libraries. Journal of Intelligent Information Systems, vol. 31, n. 1, pp. 53-84 (2008)
6. Golder S.A., Hubemann, A.: Usage Patterns of collaborative tagging systems. Information Science (2006)
7. Mishra, S., Ghosh, H.: Effective Visualization and Navigation in a Multimedia Document Collection using Ontology. Pattern Recognition and Machine Intelligence, pp. 501-506 (2009)
8. Kennedy, L., Naaman, M., Ahern, S., Nair, R., Rattenbury, T.: How flickr helps us make sense of the world: context and content in community-contributed media collections. ACM Multimedia, vol. 99, n.7, pp. 631-640 (2007)
9. Moscato, V., Penta, A., Persia, F., Picariello, A.: A System for Automatic Image Categorization. International Conference on Semantic Computing, pp. 624-629 (2009)
10. Moscato, V., Penta, A., Persia, F., Picariello, A.: A System for Building Multimedia Ontologies from Web Information Sources. IIR, pp. 89-93 (2010)
11. Paliouras, G., et al.: Bootstrapping Ontology Evolution with Multimedia Information Extraction, Knowledge-Driven Multimedia Information Extraction and Ontology Evolution. LNCS 6050, pp. 1-17 (2011)
12. Petridis, K., et al.: M-OntoMat-Annotizer: Image Annotation Linking Ontologies and Multimedia Low-Level Features. Knowledge-Based Intelligent Information and Engineering Systems (LNCS), vol. 4253, pp. 633-640 (2006)
13. Stamou, G. et al.: Multimedia annotations on the semantic Web- IEEE Multimedia, vol. 13, pp. 86-90 (2006)

Nicola Ferro and Letizia Tanca (Eds.): SEBD 2012, Edizioni Libreria Progetto, Padova, Italia
ISBN: 978-88-96477-23-6,