A Hybrid Ontology and Visual-based Retrieval Model for Cultural Heritage Multimedia Collections

Stefanos Vrochidis, Charalambos Doulaverakis, Anastasios Gounaris,
Evangelia Nidelkou, Lambros Makris, Ioannis Kompatsiaris
{stefanos, doulaver, gounaris, nidelkou, lmak, ikom}@iti.gr

Informatics and Telematics Institute
Thessaloniki, Greece

Abstract. Nowadays, an increasingly growing demand for advanced multimedia search engines is arising, as huge amounts of digital visual content are becoming available. The contribution of this paper is the introduction of a hybrid multimedia retrieval model accompanied by the presentation of a search engine that is capable of retrieving visual content from cultural heritage multimedia libraries as in three modes: (i) based on their semantic annotation with the help of an ontology; (ii) based on the visual features with a view to finding similar content; and (iii) based on the combination of these two strategies in order to produce recommendations. To achieve this, the retrieval model is composed of two different parts, a low-level visual feature analysis and retrieval and a high-level ontology infrastructure. The main novelty is the way in which these two co-operate transparently during the evaluation of a single query in a hybrid fashion, making recommendations to the user and retrieving content that is both visually and semantically similar. A search engine has been developed implementing this model which is capable of searching through digital libraries of cultural heritage collections, and indicative examples are discussed, along with insights into its performance.

1   Introduction

Multimedia content management plays a key role in modern information systems. Usually in the past, medium size, mainly textual archives used to be the only resources for knowledge management, whereas nowadays, organizations and companies handle very large quantities of multimedia information, basically due to the dramatic decrease of digital storage cost. As a result, from personal photo collections to media archives, cultural heritage collections and bio-medical applications, an extremely valuable information asset is in the form of images and video. To provide the same functionalities for the manipulation and knowledge retrieval from such visual content as those provided for text processing, a key aspect is the development of more efficient search engines for image and video files.

To date, two main approaches to image search engine techniques have been proposed, namely annotation-based and content-based. The former is based on image metadata or keywords that annotate the visual content or they refer to the properties of the image file. Examples of image file properties include the name of the image file, its

creation date, copyright information, image format, resolution and so on. On the other hand, content metadata correspond to the properties of the entities depicted.

Examples of image file properties include the name of the image file, its creation date, copyright information, image format, resolution and so on. On the other hand, content metadata correspond to the properties of the entities depicted. For instance, a photograph from someone's holidays may be tagged with concepts such as "beach" and "mountain". Several variants of annotation-based multimedia search engines have been proposed. Some of them assume manual annotation (e.g., [6]), while others provide support for automatic annotation, by exploiting, for example, relevant text associated to the image and by applying natural language processing techniques.

A well known example that falls into this category is Images Google Search[1].

This search approach has benefitted significantly from the advances in the Semantic Web and ontologies (e.g., [18]), so that annotations can have well-defined semantics. Ontologies are "an explicit specification of a conceptualization" [7], and they guarantee firstly a shared understanding of a particular domain, and secondly, a formal model that is amenable to unsupervised, machine processing. The use of ontologies has also made possible the integration of different content under a unified description base where various collections can be accessed using a common querying framework. For example, some museums use ontologies for storing and describing their collections, so that users can browse and explore the museum collections, and understand the way in which the items are described and organized, based on their semantics. Indicative examples of such systems are Artefacts Canada[2] and MuseoSuomi[3].

However, annotation and semantic-based search are often insufficient when dealing with visual content. When searching through cultural heritage material for example, it is very common to look for images that are visually similar but may be annotated in a different way (thus a strict annotation search cannot retrieve all relevant images). In addition, in many cases, manual annotation is not available or is incomplete. To tackle this problem, a second complementary approach has been devised: content-based search. The core idea is to apply image processing and feature extraction algorithms to the visual content and extract low-level visual features, such as color layout and edge histogram [19]. The retrieval is performed based on similarity metrics between such features, attempting to imitate the way humans perceive visual similarity (e.g., [14]). The most common way for conducting a search is using the query-by-example methodology [10],[15] where a user is expected to have a visual multimedia item depicting the object or scene of interest.

This paper focuses on a hybrid retrieval model by combining in an novel way the content and annotation based approaches. A search engine has been developed implementing this model using images from the culture domain.

This new model does not just provide an integration or mixing of results but uses the results from each stand-alone approach to generate new searches and enhance the final list of images presented to the user taking into account both semantic and visual similarity. A search engine has been developed implementing this model using images

---

[1] http://images.google.com/

[2] http://www.chin.gc.ca/English/Artefacts_Canada/

[3] http://www.museosuomi.fi/

from the culture domain. The ontology-based search builds upon the knowledge base storing the pre-existing semantic annotations of the collection items. Regarding content-based search, the engine employs state-of-the-art techniques, which involve automatic segmentation of 2D visual content and MPEG-7 features extraction, while the novel, hybrid search functionality is capable of extending either the content-based search by making user suggestions of additional, potentially interesting results. that are semantically related to the visually similar ones, either the semantic search by recommending images which are visually similar with the results. Extending the procedure for video files, it is assumed that temporal aspects are not important in the retrieval task, thus the search engine can analyze shot keyframes independently, which leads to a unified way of handling image and video.

The remainder of the paper is structured as follows. Section 2 introduces the retrieval model on which the engine is built and describes the visual and semantic retrieval. The implementation and evaluation procedures including the presentation of the search engine, the ontology structure, the visual content and the ground truth definition are presented in Section 3, while examples, results of the hybrid engine and insight into the performance of the different approaches appear in Section 4. Section 5 deals with the related work and eventually, section 6 concludes the paper.

2 Hybrid Retrieval of Visual Content

As mentioned previously, the search engine described hereby supports three modes of queries and retrieval of images and video, namely:

1. content-based retrieval,
2. ontology-based retrieval, and
3. hybrid retrieval, which builds upon the combination of the two aforementioned methods.

In the remainder of this section, the technical details of the three different flavors of the search engine will be discussed.

2.1 Combining Visual and Semantic Information

The main objective behind this retrieval model is to allow a user to complement a query primarily addressed using one of the visual or semantic mode with the other. With a view to supporting such functionality, the hybrid search engine provides a novel retrieval method, in which both visual and ontology search are employed for the same query. The novel retrieval method, employed by the model, automatically combines different types of search results, and complements content-based search with ontology-based search and vice versa. Starting with one mode, information arising from the complementary mode is used to enhance the results. The additional results presented are considered to be a set of recommendations for the user by broadening the desirable query and they are generated in a transparent way. It is important to note that the hybrid engine generates the new queries involved to retrieve more results in a way transparent to the user. The final result sets are integrated and are presented to the user in a unified

manner. This retrieval model is based on the combination of this two main retrieval techniques. The hybrid output can be produced either with a semantic based query either with a visual search. The produced output is considered to be a set of recommendations for the user by broadening the desirable query and illustrating results which are visually or semantically related to the initial results.

The mathematical model of the hybrid retrieval system is described below for both cases. Lets assume that the function $Sem(data_{rdf}, q_{sem})$ is producing the desired output given the data and the query based on the semantic data formed in RDF[4] language by retrieving the results from the Knowledge Base:

$$Res_{sem} = Sem(data_{rdf}, q_{sem}) \tag{1}$$

where $data_{rdf}$ are the metadata stored in RDF in the Knowledge Base and q_{sem} is the query string in RDFQL or SeRQL. Here it has to be mentioned that the Res_{sem} are ranked randomly by the function $Sem(data_{rdf}, q_{sem})$ as all the output results satisfy 100% the submitted query.

In a similar way, the function: $Vis(data_{desc}, q_{vis})$ outputs the results from content-based search using as data the extracted descriptors of the multimedia content and the proper input from the user.

$$Res_{vis} = Vis(data_{desc}, q_{vis}) \tag{2}$$

where $data_{desc}$ represent the extracted descriptors of the multimedia content and q_{vis} represents the desirable input (i.e one or a set of images) for which visually matching content expected to be retrieved and displayed. The function $Vis(data_{desc}, q_{vis})$ outputs Res_{vis} in a specific ranking based on the similarity coefficient which derives from the calculation of the distances of the extracted descriptors for the objects included in the query.

Subsequently, two cases of hybrid search are defined: (i) the visual search, where the system, given the desirable query, produces visually similar results with the initial object accompanied by a set of recommendations deriving from the transparent semantic query that the visual results produce; and (ii) the semantic search, where a user can submit a query by browsing the ontology fields and acquire the results that illustrate the content which satisfies the constraints of the query complemented by recommendations based on visual similarity of the initial results. In the case of visual search the initial visual results are processed in order to provide complementary semantic recommendations based on the semantic feature that is most commonly shared between the initial results. On the other hand, when a Semantic search occurs, the results that are produced consist of the initial semantic set accompanied by the visual one, which is based on the visual similarity of a hypothetical image averaging the MPEG-7 features of the initial results. The whole procedure is supported by a detailed mathematical model.

In visual search the output consists of two sets of results: the initial results produced by Res_{vis} and the set of recommendations Rec_{sem} based on semantic search and given by:

$$Rec_{sem} = Sem(data_{rdf}, ResToQ_{sem}(Res_{vis})) <=> \tag{3}$$

---

[4] http://www.w3.org/RDF/

$$Rec_{sem} = Sem(data_{rdf}, ResToQ_{sem}(Vis(data_{desc}, q_{vis}))) \tag{4}$$

where the function $ResToQ_{sem}$ creates a new query based on the first set of the results in order to retrieve the semantically related content. This function could be based on a combination of several algorithms in order to output a query which contains the information from the visual results. The algorithm adopted for the tests and the evaluation with the search engine for cultural content was based on the semantic concept which appeared more frequently between the results. More specifically The aforementioned function exploits the initial set of results by processing the ontology fields of every output object in order to define the semantic annotation which is mostly shared by these results. The query produced leads to a search for content that shares the specific common value in the chosen ontology fields with the results.

The final set of results Res is the set of results from visual similarity Res_{vis} enhanced by the recommendation results Rec_{sem}:

$$Res = Res_{vis} \cup Rec_{sem}. \tag{5}$$

On the other hand, when a Semantic search occurs the results that are produced consist of: the first set provided by Res_{sem} and the second set Rec_{vis} illustrating the recommendations:

$$Rec_{vis} = Sem(data_{desc}, ResToQ_{vis}(Res_{sem})) <=> \tag{6}$$

$$Rec_{vis} = Vis(data_{desc}, ResToQ_{vis}(Sem(data_{rdf}, q_{sem}))) \tag{7}$$

where the function $ResToQ_{vis}$ constructs a query taking into account the visual features of the initial results. The algorithm used for this function produces descriptors of an average hypothetical object by averaging the descriptors of the results. Assuming that the resulted objects would share common visual features due to their semantic relation, the query constructed by the function would search for objects visually similar with the hypothetical one.

The final set of results is:

$$Res = Res_{sem} \cup Rec_{vis}. \tag{8}$$

The architecture of the Hybrid retrieval model is illustrated in Figure 1. While methods for fusing together plain metadata and low-level features mainly in web-based search systems have been employed (e.g., [16], [3]), to the best of the authors' knowledge this is the first attempt to combine semantic and visual-based feature sources in order to provide a more efficient retrieval strategy.

2.2 Content-based Retrieval

In this retrieval mode, described by (2), users are able to perform a visual-based search by taking advantage of low-level multimedia content features. The retrieval system can handle 2D still image and potentially video. In this mode, the user provides, as the input query, an example of the multimedia content she or he is interested in, and, based on the extracted descriptors of the input and the indexed offline-generated descriptors of
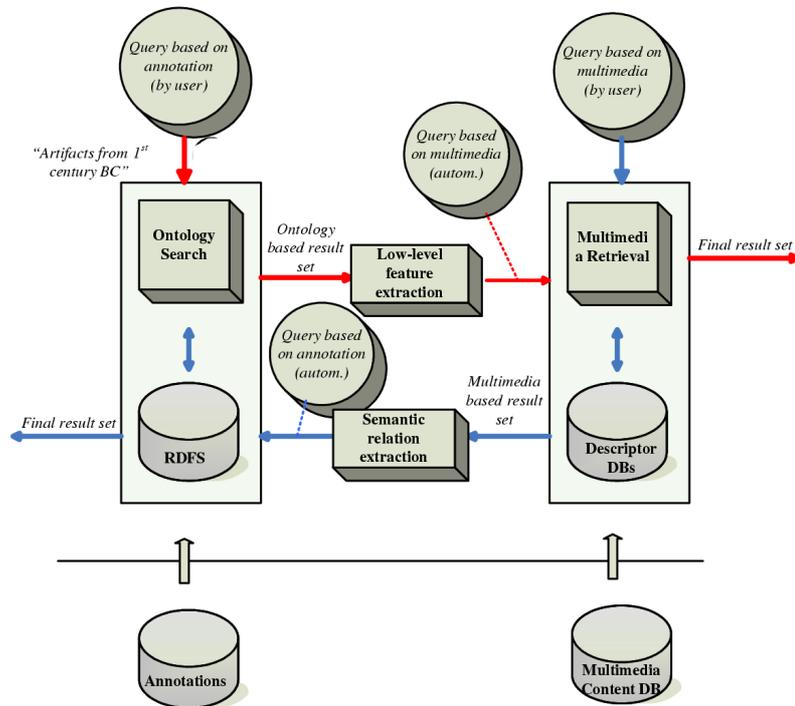
Fig. 1. Hybrid retrieval model architecture.

the content repository, the system performs a visual similarity-based search and the relevant results are retrieved. The analysis of 2D images is performed using the approach described in [13] and involves the extraction of MPEG-7 features.

For proper handling of the various content types, different strategies are employed for each type in the offline analysis process. This process involves the following steps:

Step 1: Segmentation algorithms are applied.
Step 2: Low-level visual features are firstly extracted, and secondly normalized.
Step 3: The results generated by this offline process are indexed and stored in a database.

Figure 2 illustrates the 2D image analysis with potential application to video files.

At runtime, the first two steps are followed for the input query, and the intermediate results are compared against the results generated offline in order to detect visually similar images.
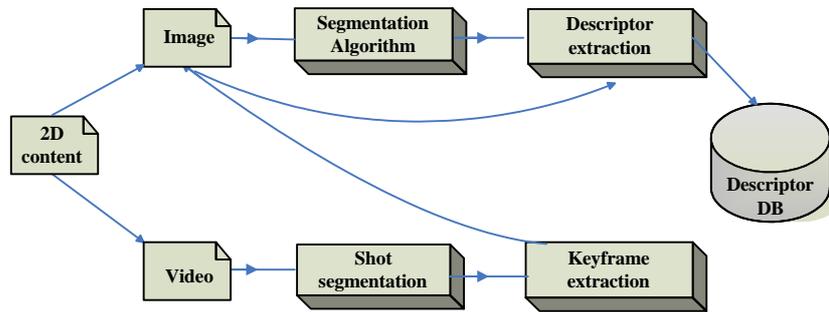
Fig. 2. The offline image and video content analysis process.

The analysis of 2D images is performed using the approach described in [13] and involves the extraction of MPEG-7 features.

More specifically, analysis of 2D images is performed in a two-step fashion. To enable meaningful region detection in the available cultural heritage images collections, a segmentation process takes place using the approach described in [13]. There are several advantages in using regions for image retrieval and these are mainly derived from the fact that users usually search for objects displayed in images rather than whole images instead. This is the typical case in the area of cultural heritage as the main interest in retrieval is the item being displayed in an image regardless of any surroundings or background. The applied color image segmentation algorithm produces connected regions that correspond to meaningful objects shown in the image by performing the segmentation process in the combined intensity and texture feature space. Results of this segmentation approach on different images from the cultural area are presented in Figure 3.

The second step in analysis involves low-level feature extraction from the resulting regions of the segmentation mask and also from the whole image itself. For this purpose, the MPEG-7 features were selected as they represent the state of the art in low-level visual descriptors. For the extraction, the MPEG-7 eXperimentation Model (MPEG-7 XM) [1] was used as it realizes the standardized descriptors and apart from extraction it also utilizes methods for similarity based retrieval.

This procedure could be extended or video analysis. In such a case the video stream is firstly divided into shots using the method described in [11]. For each detected shot, a keyframe is extracted which is treated as a compact representation of the entire shot. This keyframe is then analyzed as in the still image case, i.e., it is segmented into regions and feature extraction using MPEG-7 XM is performed.

2.3 Ontology-based Retrieval

This search mode, described by (1), is more appropriate for the cases in which the user knows to an adequate degree of confidence the semantic annotation of the material he or
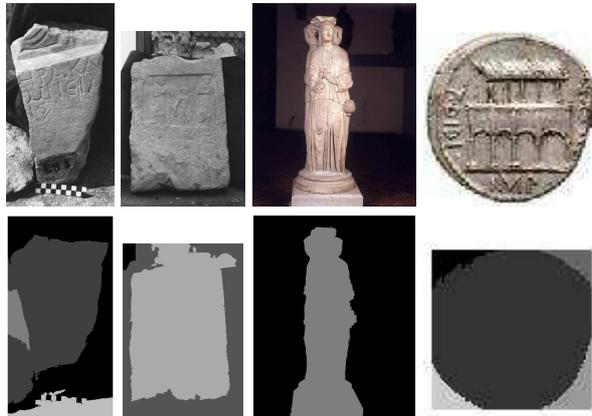
Fig. 3. Cultural heritage images (top) and their corresponding segmentations (bottom).

she is searching for and for the cases. To submit a query, the user provides constraints on the concepts of the five ontologies of Section 3.2. During search time, the system retrieves the semantically connected content according to users selections. In this way, the system can automatically retrieve items, not necessarily of the same type (e.g., coins and inscriptions), that were made, for instance, of the same material, or in the same period or were found in the same place. In this search mode, the system is capable of handling complex queries that require the combination of multiple concept-based search criteria and thus can retrieve different cultural items that share common data. Other related systems, as that of the State Hermitage Museum, do not allow for this kind of functionality but restrict user choices to a specific item category.

Figure 4 illustrates the application of this technique by presenting a proper interface for ontology based search.

In the example in the figure, a search for the available bibliographic references has been requested by selecting the appropriate class from the ontology and, as filtering predicates, the user has selected items that are exhibited in a specific museum and are referenced in a specific historical book. The results are displayed as shown in Figure 5. This is an example illustrating that the developed system can handle complex queries that require the combination of multiple concept-based search criteria and thus can retrieve different cultural items that share common data.

3 Search Engine and Evaluation Corpus

The evaluation procedure took place with the employment of a search engine based on the aforementioned retrieval model and it is capable of retrieving cultural visual content.

Fig. 4. Search engine interface where the user can select one or more ontology instances and concepts to form the search criteria. Bottom: the results of the ontology-based search for specific selections.

This engine is used to search through a collection of ancient Greek coins and inscriptions, a collection of paintings and photographs, however its architecture and search approach are actually domain independent. In order to implement the ontology-based retrieval, an ontology-based representation was created with the objective to provide enhanced, unified access to heterogeneous cultural heritage digital databases. The approach followed has been to comply and align with the CIDOC-CRM [5] core ontology, proposed as an ISO standard for cultural heritage material structure and representation, in order to guarantee the compatibility with other ontology initiatives from the same domain.

Illustrations of the developed search engine are displayed in Figure 4. The design of the GUI and the ontology browser plays a significant role for ontology-based retrieval. As shown in Figure 4, the GUI of our search engine provides a view of the ontologies, enabling the browsing through their structure and hierarchy; selected concepts of each one are automatically organized according to their class hierarchy in a tree-like fashion. Moreover, the possible values for each concept are displayed, in order to help the user to formulate the search criteria. Using this design in the system interface has proven to be more user-friendly as all the selections are readily available without introducing any confusion, while in the same time the underlying ontology structure is obvious, letting the user understand the employed knowledge structure used. This would not be feasible if a drop-down list GUI architecture had been selected instead. Apart from the ontology-based retrieval, the search engine supports content-based queries in order
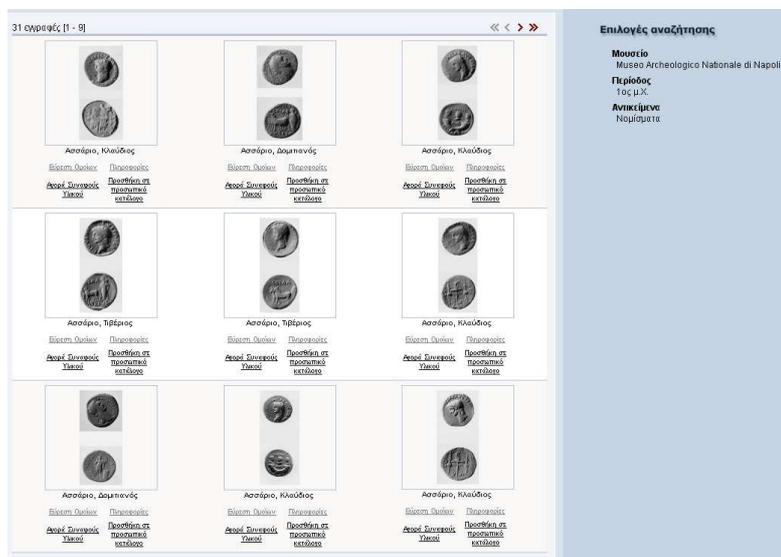
Fig. 5. The user has asked for artefacts from coins collection (using the "Artefacts" concept) displayed in the "Museo Archeologico Nationale di Napoli" (using the "museum" concept) and are dated in the 1st century BC (using the "period" concept).

to produce results depending on visual similarity. Combining the two techniques, the search engine is capable of providing the hybrid searching functionality. in order to enhance the initial results and generate recommendations for the user based on the hybrid retrieval model as was described in detail in Section 2.1.

3.1 Visual Content

The main content provider is the Center for Greek and Roman Antiquity (KERA)[5], which offers a large collection of inscriptions and coins from the Greco-Roman time period , accompanied with detailed documentation. Furthermore, a rich collection of Greek paintings from the 18th to 20th century is provided by the Greek museum: Teloglion Foundation of Art[6] while a large collection of photographs is offered by Alinari Photographic Archives[7].

In addition to the above, Alinari Photographic Archives offered a large collection of black and white photographs covering a vast variety of themes mainly from the 19th century.
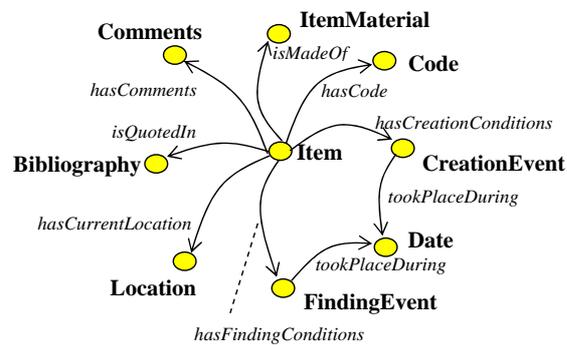
---

[5] http://www.eie.gr/nhrf/institutes/igra/index-en.html

[6] http://web.auth.gr/teloglion/

[7] http://www.alinari.com/

Fig. 6. A graphical representation of the concepts in the ontology and their relations.

3.2 Ontology

Cultural heritage collections are accompanied by a rich set of annotations that describe various details related to each item regarding historical data or details regarding administrative information like, for example, current exhibition location.

However, these annotations are often unstructured or registered in a non-standard form, usually proprietary, for every collection, which renders them unusable for inter-collection searching. To overcome this problem, appropriate ontologies for the cultural heritage domain have been defined. Such ontologies can be used for searching purposes when the search criteria are the collection item annotations rather than the visual appearance. The ontologies are also defined in a way that makes them suitable for administrative purposes of the collection material as during the design process the available metadata were studied and each ontology was defined in a way which was best to represent these data but also taking into account that they should also be usable in a search engine scenario.

Taking into account the content originally available for our use case scenario an ontology infrastructure has been defined to efficiently describe and represent all knowledge related to each collection. The proposed architecture consists of two layers and makes use of five different RDF(S) ontologies, namely *Annotation*, which is generic, and *Coins*, *Inscriptions*, *Paintings* and *Photographs* which are specific to the collection itemsets of our scenario. (Figure6). These ontologies are detailed below.

A large set of information fields inside the image annotation set is common for each item, regardless of the collection that is part of. As such, it was decided to use a separate, higher-level ontology specifically intended for representing this kind of concepts and relations, which cover information like date and place of creation, current location, construction material, dimensions, etc. Such data is an example of properties that appear in and characterize every item inside the collection. Consequently, the role of the *Annotation* ontology (Figure 6) is to conceptualize and hold all common data in a structured way, thus forming a representation standard for every collection to be integrated with the search engine. Additionally it is much easier to inter-relate the different items

Original relation	CIDOC-CRM mapping
Item $\rightarrow$ *isMadeOf* $\rightarrow$ ItemMaterial	E22.Man-Made_Object $\rightarrow$ *P45F.consists_of* E57.Material
Item $\rightarrow$ *isQuotedIn* $\rightarrow$ Bibliography	E22.Man-Made_Object $\rightarrow$ *P70B.is_documented_in* E31.Document

Table 1. Example Mappings of the main General Item Attributes defined in our ontology (left) to the CIDOC-CRM (right)

using these concepts as they are all represented in the same ontology while keeping the ontology infrastructure modular.

The properties that are specific to a collection item category are captured by complementary ontologies; more specifically there is a separate ontology for each category, as the particular details that correspond to the collection items can vary greatly for each class. For example, the information that one requires to search through Coin collections, such as monetary subdivision, is significantly different from the information used for Inscriptions searching, such as inscription text. A thorough study of the annotations has shown that this kind of specific information does not overlap across items as is the case with the *Annotation* ontology. As a result, the definition of a *Coins* and an *Inscriptions* ontology was the most appropriate approach in our case, since it can efficiently handle the data. Moreover, it does not restrict the extensibility of the system as the addition of cultural items of an additional type only requires the definition of a specific domain ontology for that type, and the mapping of its common data to the *Annotation* ontology, i.e., the architecture is extensible and in the future it can cover additional collections like statues, vases, manuscripts, and so on.

As a further step, to support interoperability of the system with other semantic-enabled cultural heritage systems, the aforementioned ontologies were mapped to the CIDOC-CRM [5] core ontology which has been proposed as an ISO standard for cultural heritage material structuring and representation. To enable this functionality, appropriate mappings between the concepts of our defined ontologies and the CRM were drawn.

This rather time consuming process gives an advantage to cultural heritage collection management systems, independently of whether they are ontology based or not, regarding the ability to allow for seamless cross-communication with other similar systems as inter-collection relations can be made public using a standardized schema. Some example mappings are presented in Table 1, where on the left column it is illustrated how concepts and relations are represented in our ontology, whereas on the right column the corresponding relation in CIDOC-CRM is shown.

To summarize, the ontology infrastructure of the search engine developed has the following key characteristics:

– It follows a modular design approach. There is a separate module for each collection itemset, which is linked to a generic module, called the *Annotation* ontology.

Apart from a clear design, this approach guarantees the extensibility support of the system.

– The ontologies are mapped to the ISO-standard CIDOC-CRM core ontology, thus allowing for interoperability with other, third-party ontologies from the cultural heritage domain.

3.3 Ground Truth Definition

The ground truth used in order to evaluate the results of the experiments were different for each retrieval mode.

Regarding the content-based experiments as ground truth was considered the (subjective) visual similarity of the objects. More specifically the visual features which were taken into account to prove visual similarity were the shape and the color while the existence of visually related regions between the objects can also be considered as factor of visual resemblance.

The results from ontology-based queries could be easily evaluated due to the existing annotations.

The recommendations, which are results of the hybrid mode, are considered to be results related to the initial set of the semantic or visually based results. The definition of the term recommendation can be subjective however it can be defined as any result which is related to the initial output in terms of visual or semantic similarity. The recommendations serve the purpose of broadening the query by presenting an additional set of results that could be of interest to the user. Hence as ground truth was considered any semantic or visual relation, which can properly justified, that the recommendation set could have with the initial set of results.

4 Results

The content-based and ontology-based modes are complementary to each other, and as such, it is meaningless to compare them directly in terms of metrics like precision and recall.

They actually cover complementary user needs, and consequently, the hybrid proposal is of particular practical interest. This section serves two purposes. Firstly, to demonstrate the advanced functionality of the hybrid search engine, through some use cases. Secondly, to provide insights into the performance of the two search flavors, mainly content-based and ontology-based. In this section the advanced functionalities of the hybrid search engine are demonstrated through use cases and insights into the performance of the two search flavors are provided.

4.1 Hybrid Search: Use Cases

The hybrid search engine is capable of detecting implicit semantic relationships between visually dissimilar images, and extract the relevant artefacts. To demonstrate this capability, two use cases are presented in this section, which are summarized in Figures 7 and 8.

In a the first use case (Figure 7), the input query is the painting "Iera Moni Agiou Pavlou" ("Holly monastery of Saint Paul"). (1933), which is created by Reggos Polykleitos and is displayed in the Teloglion Foundation of Art. During the content-based search visually similar inscriptions are extracted (Figure 7a). As shown in the figure, the results have several visual features in common, such as depiction of buildings and mountains. To fire an ontology-based search, the system retrieves the most common semantic feature, which happens to be the name of the painter. As such, the system is capable of automatically returning the same results as if the user was capable of submitting a query of the type (in natural langauge) *"Find all the artefacts that look like the painting "Iera Moni Agiou Pavlou" or all the Paintings created by Reggos Polykleitos"* (Figure 7b). The first set of results shows visually similar images by illustrating mostly paintings with monasteries and buildings while the recommendations include paintings of the same painter who has created paintings of similar themes (monastery, churches, etc.) and in addition a portrait which cannot be retrieved by visual similarity (Figure 7).

The second use case follows the opposite approach where the initial query is based on ontology fields while the set of recommendations derives from visual similarity. In this scenario, the user searches for inscriptionsfor inscriptions characterized as anaglyphs. The first set of results includes the two inscriptions which are named as anaglyphs (Figure 8a) while the recommendations provide results visually similar with the two inscriptions. As it can be observed in Figure 8b the recommendation images include more inscriptions with similar shapes and figures with the two anaglyphs in addition to some images from the photographs collection. However it is remarkable that pictures in the photographs illustrate visually similar regions with the two initial inscriptions as human shapes and sculptures. Consequently the recommendations broaden the initial query as they reveal visually similar images of potential interest that the user may not have been aware of when submitted the query.

From the aforementioned use cases, two important observations can be made. Firstly, the hybrid search engine is capable of recommending items to the user based on the semantic or the visual similarity of the results which outcomes from the user input query, without any additional effort from the user. Secondly, the user receives feedback on the search criteria of the transparent, ontology-based or visual query. This is a key aspect for the better interpretation of the additional results.

4.2 Performance Insights

In the previous section, three different search policies were presented that provide three complementary options to query formulation, so that the users can find their desired content even in the case where the search criteria are rather complex. In this section, we present a closer inspection on the efficiency and the performance of each of the two basic methods (content- and ontology-based) and draw conclusions on the advantages and disadvantages of each method with respect to precision of retrieval and response times.

The experiments were conducted on a PC, with a P5 3.0GHz Intel CPU and 1GB RAM. The knowledge base containing the ontological metadata is Sesame 1.2 running a MySQL DBMS at the back-end. MySQL is also used to store the actual non-multimedia content and the links to the multimedia files. The dataset is consisted of roughly 4000

images including inscriptions, coins, paintings and photographs along with a complete set of semantic annotations. The visual descriptors are stored in a collocated MPEG-7 XM server. For both content-based and ontology-based search, five queries, either visual or semantic, were used and the mean times are presented below. To evaluate the content-based search we selected five random images.

Figure 9 shows the Precision-Recall diagram for the content-based retrieval. The curves correspond to the mean precision value that was measured after several retrieval tasks. For the ontology-based search since it is based on selecting available concepts describing the content, the estimation of Precision-Recall diagram is not relevant. The average response time for the ontology-based search is 0.163 sec, while for the content-based search is 0.773 sec. For the content-based search it includes the communication with the MPEG-7 XM server and if we assume that the metadata used (i.e., the visual features) are evaluated and stored in the preprocessing step in the multimedia database instead of the MPEG-7 XM server, then the time cost of content-based search is reduced to 0.068 sec.

Comparing the two methods, we should keep in mind that (i) these methods work on different representations of the available data and their use is intended to satisfy different needs; and (ii) ontology-based search presupposes the manual annotation of collection items. In summary, ontology-based search aims at making use of the semantic annotations associated to an item, with respect to historical data (e.g., date of creation, place, etc.), while the content-based search aims at making use of lower level characteristics of the multimedia content corresponding to an item, like shape and color distribution, which can be automatically extracted. Such information is not likely to be found in the metadata accompanying a cultural heritage collection. Someone looking for items that are similar in shape, for instance, will use visual similarity as compared to a user interested in finding items that belong to a certain time period, and thus benefits more from the semantic search engine.

The behavior of the hybrid search is expected to combine the benefits of the other two approaches providing recommendations to the user in order to broaden the query. Precision-Recall graphs for the recommendations are not presented as these strongly depend on the nature of the retrieval task, and on the objective and purpose of the user when submitting a query. A recommendation can be considered subjective enough especially when personalization of the user is not applied. The personal background of the user is very important when speaking about cultural content as i.e an archaeologist would have different expectations from a typical visitor of a museum regarding recommendations of a search engine. Consequently, solid measurement method for the hybrid search is difficult to obtain because of the strongly subjective nature of this proposed search option. However the recommendations provided by the system were considered as satisfactory, as they are related either semantically either visually with the initial set of the results and this conclusion was supported by an adequate number of questionnaires which where filled in by users during an evaluation task of the search engine.

Concluding, hybrid search is proposed as a novel way of combining the above two methods to provide results sets that could potentially be of relevance, and are based both on visual features and on the concepts defined in the ontology.

5 Related Work

Multimedia search engines have attracted a lot of interest both from the web search engine industry (such as Google, Yahoo!, and so on) and from academia. Also, the emergence of MPEG-7 standard has played a significant role in content-based search becoming a mature technology. For a survey, the reader can refer to [2, 4]. However, to the best of the authors' knowledge, no framework has been proposed on the combined use of ontology- and content-based retrieval.

Several efforts for the efficient management and indexing of cultural heritage collections have been reported, however they all lack the integrated functionality of the presented hybrid search engine. In MuseoSuomi [9] automatic semantic association generation was used through database schemas and metadata using underlying RDF(S) ontologies. The system supports runtime distributed querying through geographically remote collections while at the same time maintaining semantic relations between collection items. The users can search through the collections using a multi-facet interface [8]. Although the system can efficiently handle semantic metadata, there is no support for visual search. The SCULPTEUR project [18] employs the CIDOC-CRM ontology to enable concept-based browsing of the annotations. Additionally cross collection searching is enabled by contacting web-services, and more specifically an extension to SRW [8], to allow for access to different content repositories. CIDOC-CRM is used for mapping each collection's proprietary schema to a standardized structured representation. Queries to the collection are in CQL (Common Query Language) and an mSpace-based [12] interface is employed to enable searching using the CRM's concepts. Multimedia search is also supported, both 2D and 3D. Nevertheless, SCULPTEUR makes no proposal for coupling semantic data and low-level visual features. The ongoing EU project eChase [17] also uses the CIDOC-CRM ontology to map various cultural heritage collections. As the mapping is left to the cultural heritage content providers, there can be different mappings for the same CRM relation thus resulting in loss of semantic links. To tackle this, a reasoner will be employed to determine semantic equivalences. The collections will be queried using the SRW extension developed by SCULPTEUR. In eChase content-based search is also planned to employed but the available demonstrator does not support such kind of functionality. STITCH, which is part of eChase, tries to extract semantic relations and links with the use of vocabularies. In our case semantic relations are created based on the underlying ontological infrastructure. Semantic links generation through the exploitation of common words and their linguistic association found in the metadata is used in [20].

Another cultural heritage collection management system is the State Hermitage museum in St. Petersburg[9] developed by IBM. The online collection includes an ontology for describing the items where the main distinguishing concept between items is item category (paintings, sculpture, furniture, etc) and they are in their turn divided into more detailed subcategories. However there is no ability for cross-category search based on concepts such as date. Content-based image search is supported through IBM's QBIC[6] but there is no support for coupling together visual- and ontology-based search.

---

[8] http://www.loc.gov/z3950/agency/zing/srw
[9] http://www.hermitagemuseum.org

6 Conclusions

In this paper, a novel retrieval model for handling visual and multimedia digital libraries is presented in an efficient and effective manner. the retrieval model proposed model adopts three methods for retrieval: two autonomous and one combinational. the ontology-based method makes use of the formal, logic-based representation of semantic mark-up metadata accompanying each collection, while an illustrative user interface is used for graphical query formulation. This method is appropriate when the user is interested in semantically similar results. The content-based method makes use of the low-level visual characteristics of the multimedia material, in order to retrieve items with similar appearance. Although the search engine dealt with 2D images of cultural heritage content, there is the potential of extension based on the proposed model to include video content. A notable feature of this work is its modular and extensible ontology infrastructure, which provides mappings to CIDOC-CRM in order to gain interoperability with other ontologies from the cultural domain. The hybrid method, which is the main contribution of this work, makes a combined use of the previous two methods. Thus is capable of offering, as a recommendation, a more complete result set to the user, which comprises both visually and semantically similar items, while the input query remains either solely ontology-based or content-based. In a future work it could be very interesting to introduce personalization features in order to be able to evaluate in a more proper way the recommendations provided by the system. The search engine is being used for searching through cultural heritage material.

7 Acknowledgments

References

1. MPEG-7 eXperimentation Model http://www.lis.ei.tum.de/research/bv/topics/mmdb/e_mpeg7.html
2. G. Auffret, J. Foote, C.-S. Li, B. Shahraray, T. Syeda-Mahmood, and H. Zhang. Multimedia access and retrieval (panel session): the state of the art and future directions. In S.-F. Chang, editor, *MULTIMEDIA '99: Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, pages 443–445, New York, NY, USA, 1999. ACM Press.
3. D. Cai, X. He, Z. Li, W.-Y. Ma, and J.-R. Wen. Hierarchical Clustering of WWW Image Search Results Using Visual, Textual and Link Analysis. In *12th ACM International Conference on Multimedia*, October, 2004.
4. S.-F. Chang, Q. Huang, T. Huang, A. Puri, and B. Shahraray. Multimedia search and retrieval. In A. Puri and T. Chen, editors, *Advances in Multimedia: Systems, Standards, and Networks*. New York: Marcel Dekker, 1999.
5. M. Doerr. The CIDOC-CRM An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine*, 24(3):75–92, Fall 2003.

6. M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content: The QBIC system. *IEEE Computer*, 28(9):23–32, 1995.

7. T. Gruber. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5:199–220, 1993.

8. M. Hearst, A. Elliot, J. English, R. Sinha, K. Swearingen, and K.-P. Lee. Finding the flow in web site search. *CACM*, 45(9), 2002.

9. E. Hyvonen, M. Junnila, S. Kettula, E. Makela, S. Saarela, M. Salminen, A. Sreeni, A. Valo, and K. Viljanen. Finnish Museums on the Semantic Web. User's Perspective on MuseumFinland. In *Museums and the Web 2004, Virginia USA*, March 2004.

10. E. Izquierdo, J. Casas, R. Leonardi, P. Migliorati, N. O'Connor, I. Kompatsiaris, and M. G. Strintzis. Advanced Content-Based Semantic Scene Analysis and Information Retrieval: The Schema Project. In *Proc. Workshop on Image Analysis For Multimedia Interactive Services, London, UK*, 2003.

11. V. Kobla, D. Doermann, and K. Lin. Archiving, indexing, and retrieval of video in the compressed domain. In *Proc. SPIE Conference on Multimedia Storage and Archiving Systems, vol. 2916*, pages 78–89, 1996.

12. m.c. schraefel, M. Karam, and Z. S. mSpace: Interaction design for user-determined, adaptable domain exploration in hypermedia. In *Workshop on Adaptive Hypermedia and Adaptive Web Based Systems*, 2003.

13. V. Mezaris, I. Kompatsiaris, and M. G. Strintzis. Still Image Segmentation Tools for Object-based Multimedia Applications. *International Journal of Pattern Recognition and Artificial Intelligence*, 18(4):701–725, June 2004.

14. O. Mich, R. Brunelli, and C. Modena. A survey on video indexing. *Journal of Visual Communications and Image Representation*, 10:78–112, 1999.

15. M. Naphade, M. Yeung, and B. Yeo. A novel scheme for fast and efficient video sequence matching using compact signatures. In *Proc. SPIE Storage and Retrieval for Multimedia Databases*, volume 3972, pages 564–572, 2000.

16. E. Petrakis, K. Kontis, E. Voutsakis, and E. Milios. Relevance Feedback Methods for Logo and Trademark Image Retrieval on the Web. In *ACM Symposium on Applied Computing (SAC'2006), Special Track on Information Access and Retrieval (IAR)*, April, 2006.

17. P. Sinclair, M. Addis, D. Fina, P. Lewis, K. Martinez, D. Prideaux, and G. D. Bormida. eCHASE: Sustainable Exploitation of electronic Cultural Heritage. In *European Conference on the Integration of Knowledge, Semantic and Digital Media Technologies (EWIMT)*, November, 2005.

18. P. Sinclair, S. Goodall, P. Lewis, K. Martinez, and M. Addis. Concept browsing for multimedia retrieval in the SCULPTEUR project. In *Multimedia and the Semantic Web, held as part of the 2nd European Semantic Web Conference, Heraklion, Crete, Greece*, May 2005.

19. A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

20. P. Stuer, R. Meersman, and S. D. Bruyne. The HyperMuseum Theme Generator System: Ontology-based Internet Support for the Active Use of Digital Museum Data for Teaching and Presentation. In *Museums and the Web 2001: Selected Papers. Archieves and Museum Informatics*, 2001.
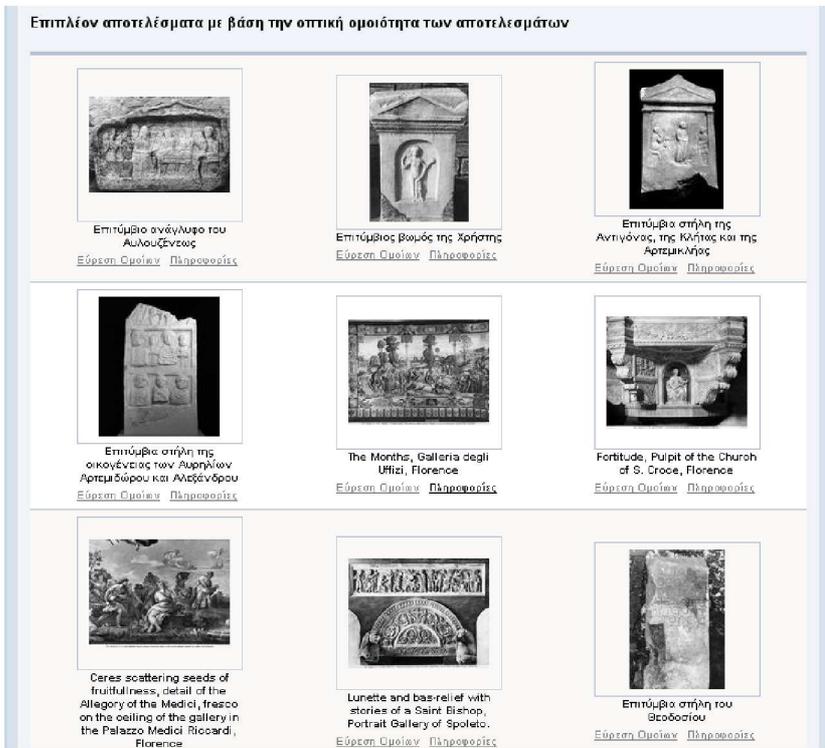
a)
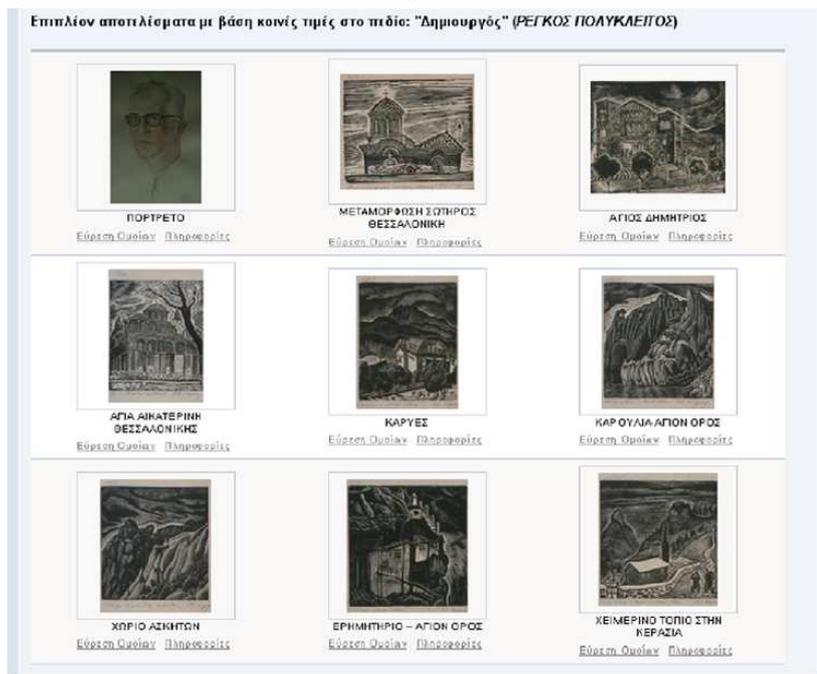
b)

Fig. 7. The first use case: (a) initial set of results derived from visual similarity search, (b)set of recommendations based on a complementary semantic query

a)

b)

Fig. 8. The second use case: (a) initial set of results based on a semantic query, (b) set of recommendations includes visually similar images
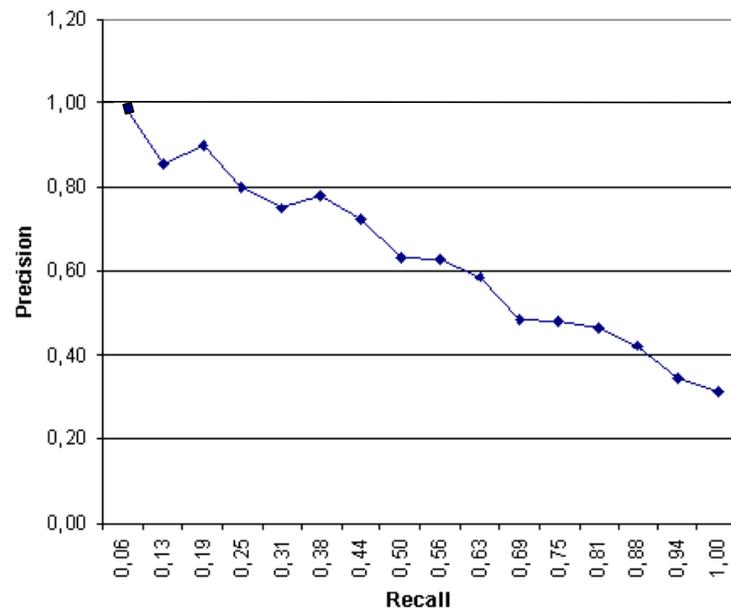
Fig. 9. Precision-Recall diagram for the content-based method.