Semantic Query Answering in Digital Libraries

Ilianna Kollia, Kostas Rapantzikos, Giorgos Stamou, and Andreas Stafylopatis

School of Electrical and Computer Engineering, National Technical University of Athens, Zographou Campus 15780, Athens, Greece

Abstract. A large activity for digitization, access and preservation of cultural heritage is taking place in Europe and the United States, which involves all types of cultural institutions, i.e., galleries, libraries, museums, archives and all types of cultural content. Semantic interoperability is a key issue in these developments. Content metadata constitute the main features of cultural items that are analysed and used to interpret users' queries, so that the most appropriate content is presented to the users. This paper presents a new semantic search methodology, including a query answering mechanism which meets the semantics of users' queries and enriches the answers by exploiting appropriate visual features, through an interweaved knowledge and machine learning based approach. An experimental study is presented, using content from the Europeana digital library, illustrating the improved performance of the proposed semantic search approach.

1 Introduction

Digital evolution of the Cultural Heritage Field has grown rapidly in the last few years. Following the early developments at European level and the Lund principles[1], massive digitisation and annotation activities have been taking place all over Europe and the United States. The creation and evolution of Europeana[2], as a unique point of access to European Cultural Heritage, has been one of the major achievements in this procedure. More than 19 million objects, expressing the European cultural richness, are currently accessible through Europeana, aiming at including all European masterpieces.

Due to the diversity of content types and of metadata schemas used to annotate the content, semantic interoperability has been identified as a key issue during the last few years. The main approach to interoperability of cultural content metadata has been the usage of well-known standards in the specific museum, archive and library sectors (Dublin Core, Cidoc-CRM, LIDO, EAD, METS) and their mapping to a common data model used - at the Europeana level: European Semantic Element (ESE, 2008), European Data Model (EDM, 2010) - to provide unified access to the centrally accessed, distributed all over Europe, cultural content. Semantic search targets on answering user queries, or enriching content providers' metadata by exploiting both explicit and implicit related knowledge. Reasoning on available knowledge bases, based on appropriate representations and languages, such as description logics, Resource Description Framework (RDF), Web Ontology Language (OWL) has been identified as the means to

[1] `http://www.cordis.europa.eu/pub/ist/docs/digicult/lund`
[2] `http://www.europeana.eu`

I. Maglogiannis, V. Plagianakos, and I. Vlahavas (Eds.): SETN 2012, LNAI 7297, pp. 17–24, 2012.

move ahead in this direction [4]. The creation of linked data stores from digital cultural heritage resources enables the linking of multiple data, assisting efficiency by permitting combined or linked searches. Semantic search and linking of data can enrich the actual content and make it useful in wider environments and contexts.

In this paper, we exploit semantic metadata representations using knowledge-based reasoning. Moreover, we show that improved query answering results can be obtained if we interweave semantic technologies with machine learning paradigms applied to appropriate features extracted from the images of the Digital Library items. Advanced semantic search based on query answering is presented in Section 2. Section 3 presents a scheme in which the knowledge used in the semantic search is interweaved with machine learning, the latter operating on features extracted from the cultural images included in the items of the Digital Library. The extraction of image features is described in Section 4. An evaluation study illustrating the performance of the proposed approach is presented in Section 5. Conclusions are given in Section 6.

2 Semantic Search in Digital Libraries

Semantic query answering refers to construction of answers to queries posed by users, based not only on string matching over data that are stored in databases, but also on the implicit meaning that can be found by reasoning based on detailed domain terminological knowledge [2]. The key is to semantically connect metadata with ontological domain knowledge through appropriate mappings. The representation formalism used for the terminological descriptions is OWL 2 (the W3C Standard for Ontology representations on the web) [4]. The theoretical framework underpinning the OWL 2 ontology representation language is *Description Logics* (DL) [1]. The DL language underpinning OWL 2 is $\mathcal{SROIQ}$. The building blocks of DL knowledge bases are atomic concepts, atomic roles and individuals that are elements of the denumerable, disjoint sets $\mathbf{C}, \mathbf{R}, \mathbf{I}$, respectively. A DL knowledge base (KB) is denoted by $\mathcal{K}=\langle\mathcal{T},\mathcal{A}\rangle$, where $\mathcal{T}$ is the terminology (usually called TBox) representing the entities of the domain and $\mathcal{A}$ is the assertional knowledge (usually called ABox) describing the objects of the world in terms of the above entities. Formally, $\mathcal{T}$ is a set of terminological axioms of the form $C_1 \sqsubseteq C_2$ or $R_1 \sqsubseteq R_2$, where C_1, C_2 are $\mathcal{SROIQ}$-concept descriptions and R_1, R_2 are $\mathcal{SROIQ}$-role descriptions. $\mathcal{SROIQ}$-concept expressivity employs conjunction ($C_1 \sqcap C_2$), disjunction ($C_1 \sqcup C_2$), universal and existential quantification ($\forall R.C$, $\exists R.C$), qualified number restrictions ($\geq R.C$, $\leq R.C$) and nominals ($\{a\}$), while $\mathcal{SROIQ}$-role expressivity allows for the definition of role inverse (R^-) and role compositions ($R_1 \circ R_2$) in the left part of the role inclusion axioms. The TBox $\mathcal{T}$ describes the restrictions of the modeled domain. The ABox $\mathcal{A}$ is a finite set of *assertions* of the form $A(a)$ or $R(a,b)$, where $a, b \in \mathbf{I}$, $A \in \mathbf{C}$ and $R \in \mathbf{R}$. An interpretation $\mathcal{I}$ maps concepts to subsets of the object domain, roles to pairs of elements from the object domain and individuals to elements of the object domain. For an interpretation to be a model several conditions have to be satisfied [1]. If an axiom ax is satisfied in every model of a knowledge base $\mathcal{K}$ we say that $\mathcal{K}$ entails ax, written $\mathcal{K} \models ax$.

We next consider concept based queries. A concept based query q is of the form $q : Q(x) \leftarrow \bigwedge_{i=1}^{n} C_i(x)$, where x is a variable and $C_i(x)$ are predicates-concept atoms.

Algorithm 1. Query Evaluation Procedure using the high level knowledge

Input: $\mathcal{K}\ \langle\mathcal{T},\mathcal{A}\rangle$: the $\mathcal{SROIQ}$ knowledge base
q: a concept based query
Output: Ans: the set of answers to the query q
$Ans := \emptyset$
$C_1, \ldots, C_n$:=$queryAtomsOf(q)$
for j=1, . . . , n **do**
for all *individual* $a \in \mathcal{A}$ **do**
if $\mathcal{K} \models C_j(a)$ **then**
$Ans := Ans \cup a$
else
if $a \in Ans$ **then**
$Ans := Ans \setminus a$
end if
end if
end for
end for
return Ans

An individual a is an answer/instance of a concept based query q posed over the DL knowledge base $\mathcal{K}$ iff $\mathcal{K} \models C_i(a)$ for i=1,. . .,n. The procedure we follow to find the answers to concept based queries is shown in Algorithm 1. The algorithm takes as input a knowledge base $\mathcal{K}$ and a query q and returns the individuals of the knowledge base that satisfy the query. This is done by iterating over the concept atoms C_j of q and over the individuals a appearing in the knowledge base $\mathcal{K}$ and by checking whether $\mathcal{K}$ entails that a is an instance of C_i. If the instantiated concept atom is entailed we add the individual to the set of answers Ans else, if it is not, we have to check whether a is already contained in Ans. In this case we remove it from the set or else we leave the set as it is. More information about the use of semantic techniques in the cultural heritage domain can be found in [6].

3 Interweaving Semantics with Machine Learning

In Section 2 we used ontologies in order to classify cultural objects to various categories and then used queries for semantic search of cultural heritage content. The creation of global axioms that hold over all items of a digital library, such as Europeana, is however, very difficult. One approach to deal with this is to use only axioms containing constraints that are known to hold over all data and leave out of the knowledge base any constraint that holds over most (but not all) of the data. In any case, the inherent, or resulting, incompleteness of the knowledge bases poses limitations to their usage in answering queries over cultural heritage content. In the current section we show that this problem can be partially overcome by interweaving the knowledge base with feature based image representations of the Digital Library items appropriately exploited by machine learning techniques; the target being to improve semantic search of cultural heritage content. The specific features of cultural heritage images that we use are presented in the next sections.

Algorithm 2. Query Evaluation Procedure using the low level visual features

Input: $trainedSVM$: a vector of trained SVM
$data$: the queried data
$features$: the features of the $data$
q: a concept based query
Output: Ans: the set of answers to query q
$Ans := \emptyset$
$C_1, \ldots, C_n$:=$queryAtomsOf(q)$
for j=1, ..., n **do**
for all $a \in data$ **do**
$output(a) := SVMpredict(features(a), trainedSVM_j)$
if $output(a) = 1$ **then**
$Ans := Ans \cup a$
else
if $output(a) = 0$ **then**
if $a \in Ans$ **then**
$Ans := Ans \setminus a$
end if
end if
end if
end for
end for
return Ans

The scope of using machine learning techniques in our method is their ability to learn from examples, in particular from the extracted features, to classify the cultural items in various concepts that can appear in queries. As a consequence, these techniques can determine the items that satisfy the corresponding query concept atoms, irrespectively of whether these items have been identified by the knowledge based component of our approach. This results in bridging the gap between restrictions imposed by ontologies and actual restrictions (visual features) that each cultural heritage item possesses. Support Vector Machines (SVMs) constitute a well known method which is based on kernel functions to efficiently induce classifiers that work by mapping the image features, and the corresponding items, onto an embedding space, where they can be discriminated by means of a linear classifier. As such, they can be used for effectively exploiting the extracted features and classify the cultural items in the different concept categories that are included in the formal knowledge. The kernel used to encode the visual knowledge through similarity between different images with respect to low level features, i.e., the feature vectors/values, is a normalized linear kernel defined as follows:

$$k_l(x, y) := \frac{x^T y + c}{||x||||y||} \tag{1}$$

where x, y are vectors of features, $|| \cdot ||$ is the Euclidean norm and c is considered zero.

Furthermore, it is possible to extend the SVM kernel so as to include individuals within ontologies [7,5]. The extension comes from a family of kernel functions defined as $k_p^F : Ind(A) \times Ind(A) \rightarrow [0, 1]$, for a knowledge base $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$. $Ind(A)$ indicates the set of individuals appearing in A, and $F = \{F_1, F_2, \ldots, F_m\}$ is a set of concept descriptions. These functions are defined as the L_p mean of the, say m, simple concept kernel functions κ_i , $i = 1, \ldots, m$, where, for every two individuals a,b, and $p > 0$,

$$\kappa_i(a,b) = \begin{cases} 1 & \begin{array}{l}(F_i(a) \in A \wedge F_i(b) \in A) \vee \\ (\neg F_i(a) \in A \wedge \neg F_i(b) \in A);\end{array} \\ 0 & \begin{array}{l}(F_i(a) \in A \wedge \neg F_i(b) \in A) \vee \\ (\neg F_i(a) \in A \wedge F_i(b) \in A);\end{array} \\ \frac{1}{2} & \text{otherwise.} \end{cases} \tag{2}$$

$$\forall a, b \in Ind(A) \quad k_p^F(a,b) := \left[\sum_{i=1}^{m} \left|\frac{\kappa_i(a,b)^p}{m}\right|\right]^{1/p} \tag{3}$$

The above kernel encodes the formal knowledge for the problem under analysis through the similarity of pairs of individuals with respect to high level features, i.e. concepts of the knowledge base. The rationale of these kernels is that similarity between items is determined by their similarity with respect to each concept F_i , i.e., if two items are instances of the concept or of its negation. A value of p = 1 is generally used for implementing (3). The extension we can use is a combined SVM kernel, computed as the mean value of the above described two kernels, i.e., $k_c(a,b) = k_p^F(a,b) + k_l(a,b)$ where k_p^F is the above knowledge driven kernel and k_l is the normalized linear kernel.

Let us now describe the way that queries are evaluated using SVMs that have already been trained to classify cultural items to concepts. Algorithm 2 shows the procedure. The algorithm takes as input the data we want to query together with their visual features and uses trained SVMs to check which items simultaneously belong to all concepts appearing in the query. $SVMpredict$ predicts the label of an item w.r.t. a concept C_j using the SVM trained to classify items to this concept. The interweaving of the two approaches, the knowledge based and the kernel based, is done by first performing Algorithms 1 and 2 for extracting the query answers from the two methods (Ans_{KB} and Ans_{SVM}) and by then disjuncting these two answer sets ($Ans = Ans_{KB} \cup Ans_{SVM}$). The enriched, in this way, results are then presented to the user.

4 Image Features Exploited by the Machine Learning Approach

The MPEG-7 standard [10] that focuses on the description of multimedia content provides among others, a set of low-level descriptors useful for tasks such as image classification, high-level concept detection and image/video retrieval. It is these features that we use as global image descriptors, i.e., *Dominant Color Descriptor* (DCD), *Color Structure Descriptor* (CSD), *Color Layout Descriptor* (CLD), *Scalable Color Descriptor* (SCD) and *Homogeneous Texture Descriptor* (HTD) and *Edge Histogram Descriptor* (EHD).

Global features do not capture local deformations and therefore are not robust to change of viewpoint or to affine transformations. For this reason we also focus on local features that are located on *interest points* of the image, like corners, blobs, salient edges, and describe their surrounding neighborhood by compact histograms (e.g. histogram of local gradients). Such points are nearly invariant to various image transformations, illumination and appearance. In particular, we use SURF (Speeded-Up Robust Features) [9] features to capture the visual properties of the images. These features have been proven to achieve high repeatability and distinctiveness. The SURF descriptor captures the intensity content distribution around the points of interest in the image.

Then, based on the extracted features, we create a visual vocabulary and we exploit it to represent the images. To understand the notion of a visual vocabulary, one should consider it as an equivalent to a typical language vocabulary, with an image corresponding to a text document. In the same way that text may be decomposed to a set of words, an image can also be decomposed to a set of *visual* words. Then, in order to compare two images, their corresponding visual words may be compared instead. To do so, we create clusters in the space of descriptors and assign each feature to the closest centroid (i.e. visual word). We do this through a fast variant of the k-means algorithm that uses approximate nearest neighbor search, in which nearest cluster centers at each iteration are assigned using randomized kd-trees. We use the FLANN [11] both in vocabulary creation and to assign visual words to image features. A histogram of constant-length can be constructed for each image, containing the appearance frequencies per visual word. This is called the *Bag-of-words* vector of the image, and it is a N_{vw}-dimensional vector, where N_{vw} is the size of the visual vocabulary. Specifically, the bag-of-words vector $bow(I)$ of an image I is a vector with elements $[tf_I(0), tf_I(1), \dots, tf_I(N_{vw})]$ where $tf_I(i)$ denotes the number of times that the visual word i was selected as a nearest neighbor of one of the interest points extracted from image I.

5 Evaluation Study

The experimental study presented in this section aims at illustrating the performance of the advanced search which exploits both semantic and visual aspects of cultural content that is accessible through the Europeana portal. We focus on the Hellenic content, which consists of 40.000 items, since for this content we also possess thematic knowledge. The thematic ontology that was manually created contains 55 categories of cultural objects (such as pottery, jewelry, stamps, wall paintings, engravings, coins) and more than 300 types. The TBox of the used knowledge base consists of the EDM together with the thematic ontology. The ABox consists of the EDM instances each one of which is described by its type, its creation date, its material, the museum it can be found at. Apart from the metadata, the visual features of the cultural objects were extracted according to the methodology presented in Section 4. Figure 1 (middle) presents images of items belonging to the 'brooch' category, that are made either of gold or copper. The binary masks of the foreground objects are also computed and used to extract color descriptors from the corresponding regions, while discarding the background. Items made of copper share similar color distributions, with these distributions being different from the item made of gold, as shown in Figure 1 (right). Based on this, the SVM classifier can separate the different categories of brooches.

In the following we apply the techniques described in Sections 2 and 3 for semantic query answering and its interweaving with the extracted visual information on the

Table 1. Accuracy (%) of Query Answering

Query	Accuracy(Algorithm 1)	Accuracy(Algorithm 2)	Accuracy(Combination)
1. $Q(x) \leftarrow OpenVase(x)$	85.5	78.4	96.9

above mentioned Europeana items. In particular we use the HermiT reasoner [3,2] and the LIBSVM library[3]. To achieve semantic search, we use the thematic knowledge for Hellenic monuments, particularly for vases (for which metadata and images are provided) that has been created in the framework of the Polemon and 'Digitalisation of the Collections of Movable Monuments of the Hellenic Ministry of Culture' Projects of the Directorate of the National Archive of Monuments[4] and which has been included in the Polydefkis terminology Thesaurus of Archaeological Collections and Monuments [8]. The knowledge used contains axioms about vases in ancient Greece, i.e., class hierarchy axioms referring to the different types of vases, such as amphora, alabaster, crater, as well as axioms regarding the appearance, usage, creation period and the material vases were made of. Using the resulting ontologies and data sets, we applied the proposed methodology to generate queries and provide semantic answers to them.

Let us assume that the user addresses a query asking for 'open vases'. Table 1 reports the accuracy of the query answering task when Algorithm 1 of Section 2 is used, based on the above mentioned terminological knowledge and instance data. It also presents the accuracy of query answering when we use Algorithm 2 of Section 3. In this case we first train SVMs (one for each concept of the query) using the extracted visual features of the items which are returned as query answers by the knowledge base. In particular, the SVMs are trained using the normalized linear kernel described in Section 3, based on the visual features and the annotated labels of the images. After that, we test the remaining data, which - erroneously - have or have not been returned as query answers based on the knowledge, using the trained SVMs according to Algorithm 2. The above means that column 3 shows the percentage of the data that the SVM 'correctly' predicts as query answers among those that were not predicted as such by the knowledge. Table 1 also shows the accuracy of query answering when we combine the results of the knowledge based

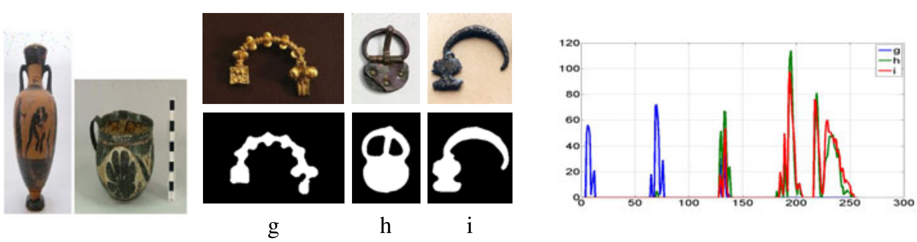

Fig. 1. A close and an open vase (left), Sample images and corresponding segmentation masks. g: made of gold, h-i: made of copper (middle), Histogram comparison for CST MPEG descriptor extracted from the regions defined by the shown binary masks (right).

[3] `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`

[4] `http://nam.culture.gr`

and visual kernel based approaches. We see that the accuracy in this case, computed as the number of generated query answers that are true over all test data, has been significantly increased. This illustrates the improved performance of the proposed advanced semantic search approach.

6 Conclusions

The current paper presents a new semantic search methodology, including a query answering mechanism that can meet the semantics of users' queries and enrich these answers by exploiting appropriate visual features, both local and MPEG-7, through an extended knowledge and machine learning based approach. We have applied this approach in the framework of a large content base (about 4 million objects) that our team – through our MINT mapping tool – has injected into Europeana. Applying this approach in a variety of subareas, such as archaeology, photography, modern arts, fashion, where specific thematic knowledge can be derived and used, as well as combining it with the evolving field of linked open data for cultural heritage are future extensions of the presented work.

Acknowledgment. The authors wish to thank the Hellenic Ministry of Culture and Tourism for providing the cultural content of the www.collections.culture.gr.

References

1. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P.F.: The Description Logic Handbook: Theory, Implementation, Applications. Cambridge University Press (2007)
2. Kollia, I., Glimm, B., Horrocks, I.: SPARQL Query Answering over OWL Ontologies. In: Antoniou, G., Grobelnik, M., Simperl, E., Parsia, B., Plexousakis, D., De Leenheer, P., Pan, J. (eds.) ESWC 2011, Part I. LNCS, vol. 6643, pp. 382–396. Springer, Heidelberg (2011)
3. Shearer, R., Motik, B., Horrocks, I.: HermiT: A Highly-Efficient OWL Reasoner. In: Proc. of the 5th Int. Workshop on OWL: Experiences and Directions (2008)
4. Motik, B., Patel-Schneider, P., Parsia, B.: OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax. W3C Recommendation (2009)
5. Kollia, I., Simou, N., Stafylopatis, A., Kollias, S.: Semantic Image Analysis using a Symbolic Neural Architecture. Journal of Image Analysis and Stereology (2010)
6. Kollia, I., Tzouvaras, V., Drosopoulos, N., Stamou, G.: A Systemic Approach for Effective Semantic Access to Cultural Content. Semantic Web Journal (2012)
7. Fanizzi, N., d'Amato, C., Esposit, F.: Statistical Learning for Inductive Query Answering on OWL Ontologies. In: Intl. Semantic Web Conference, Karlsruhe, Germany (2008)
8. Kalomirakis, D.: Polydefkis: A Terminology Thesauri for Monuments. In: Tsipopoulou, M. (ed.) Proc. of Digital Heritage in the New Knowledge Environment, Athens (2008)
9. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded Up Robust Features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
10. Chang, S., Sikora, T., Puri, A.: Overview of the MPEG-7 Standard. IEEE Trans. on Circuits and Systems for Video Technology 11, 688–695 (2001)
11. Muja, M., Lowe, D.G.: Fast approximate nearest neighbors with automatic algorithm configuration. In: Intl. Conf. on Computer Vision Theory and Application, pp. 331–340. INSTICC Press (2009)