An Ontology Construction Approach for retrieval of the Museum Artifacts Using Protégé

Manoj Kumar Sharma[1], Tanveer J Siddiqui[2]

[1] **Electronics and Communication Department, University of Allahabad
Allahabad,211002,India**

[2] **Electronics and Communication Department, University of Allahabad
Allahabad,211002,India**

Abstract

CBIR is the dominant approach in image retrieval. However, it suffers from the well-known problem of 'semantic gap'. In order to overcome this limitation we propose the use of ontology. In this paper, we implement an ontology for the domain of Museum artifact to support semantic retrieval of museum artifacts. Protégé version 4.0 is used for ontology creation and generation of Resource Descriptor Framework (RDF) schemas. The ontology is represented in RDF form. The retrieval process is illustrated with the help of examples.
.

Keywords: *Ontology, Protégé, Semantic Web, Visual Information, Information Retrieval, RDF Schemas, XML Scripts*

1. Introduction

The rapid growth in the volume of visual information makes the task of finding and accessing visual information of interest overwhelming for users. In image retrieval, two important methods are Text Based Image Retrieval (TBIR) and Content Based Image Retrieval (CBIR).TBIR uses textual keywords to describe an image. The keywords describe filename, place, creator, date the content of image. However, it is difficult to annotate each image manually. CBIR performs retrieval based on visual similarity. It is an efficient technique that uses low-level features such as color, texture and shape of an image [2]. However, there are often considerable differences between user's high level interpretation of the visual information and low level visual feature. A number of techniques have been used to overcome this limitation. This includes automatic annotation, use of semantic templates [4], and use of ontology [5], etc. In this paper, we report our attempts to use domain ontology to support museum information retrieval.

Ontology as a basis for the sharing of knowledge has been widely used in Information science and Artificial Intelligence. The knowledge-based approach is used to provide explicit domain oriented semantics in terms of defined concepts and their relationships that can be processed automatically.

Ontology defines domain concepts and relationship between them and thus provides a hierarchical structure that is meaningful to both human and machines.

In this paper, an ontology is being defined for museum domain. A museum preserves artifacts of scientific, artistic, cultural, or historical importance which attracts general public as well researcher and experts. A museum is not just a collection of artefacts but augments the presentation with useful textual description. Artefacts are organized in gallery. Ontology has four layers called, the domain layer, category layer, class layer and instance layer. The domain layer denotes the Artefacts domain name of ontology, and comprises various categories defined by the domain expert. The category layer has many category termed category 1, category 2……and category k.Each concept in the class layer contains a concept name, an attribute set and an operation set for an application domain. The domain ontology has two relationships, namely association and instance of. The concepts from this ontology can be used to annotate web resources. The Web Ontology Language (OWL) is widely accepted as the standard language for sharing semantic web content and is implemented using Protégé OWL. Protégé is an ontology development environment with a large community of active users [1]. Protégé has been extended with support for OWL, and has become one of leading OWL tools.

Our research presents the ontology design and implementation of an efficient Museum artefacts retrieval for Allahabad Museum of Cultural heritage. The goal of this paper is to define an ontology for museum domain to support semantic retrieval of museum artefacts.We discuss the classification of domain objects and define the classes, properties, and the reasoning process and generate RDF schemas and XML scripts. The retrieval process is elaborated with the help of examples

2. Ontology Construction Using Protégé

Major Building of ontology is a part of a knowledge representation process. As such, it relies on common understanding of how people represent, understand and acquire knowledge. In order to manipulate facts and ideas, people tend to impose a structure on their knowledge, similar things are grouped together according to common attributes or characteristics which they possess and then used to describe the whole group.
The process of developing ontology is called ontologization. The goal of ontologization is to construct an information structure with only the necessary information in a compact form where all intended user groups can find every bit of information as quickly and as easily as possible. An ontology consists of concepts and their relationships. To produce a formal ontology, an ontology representation language is required. We have used web ontology language (OWL) for representing concepts and their relationship hierarchy. OWL is an international standard for encoding and exchanging ontologies. OWL defines classes, properties and instances [6]. Classes define names of the relevant domain concepts and their properties. OWL ontology concepts can have references to concepts in other ontologies.This is achieved through OWL's import statement.

2.1 Classes

The important view in the Protégé OWL plug-in is the OWL classes. Classes describe concepts in the domain. The thing tab in Protégé main window displays the hierarchy of the ontology's classes. The upper region of the class is shown in a form in the center. This form allows users to edit class metadata such as name, comments, and label in multiple language. The widget in the right area of the form allows users to assign values for properties and description to class.
Annotation properties can be used to add metadata to classes.Ontologies can define their own annotation or reuse existing ones such as those from Dublin core ontology [1].In contrast to other properties annotation properties do not have any formal meaning for external OWL components like reasoner. But they are extremely important vehicle for maintaining domain information. A typical use for annotation property in museum artifact is to describe functionality of each class (concept).
In this paper, many classes and subclasses have been created under the field of artefacts. It has a number of subclasses such as Nehru Gallery, Gandhi Gallery, manmade objects, natural objects, etc. The subclasses correspond to the main categories found in museum. The class natural object is further divided into 4 subclasses. Fig. 1 depicts a part of hierarchy of domain concept.

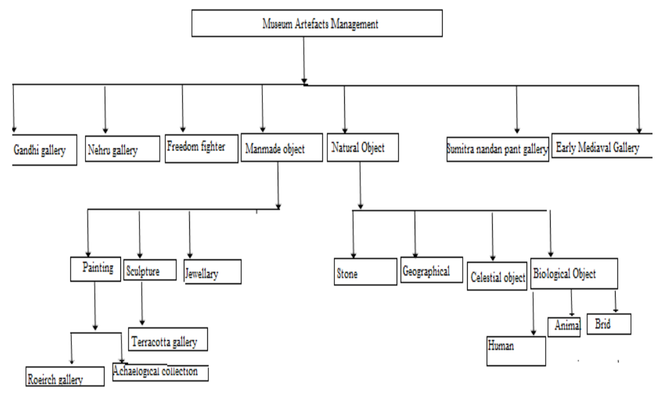

Figure 1: Concept hierarchy

The editing of classes is carried out using the *thing* tab shown in figure 2. Fig 2 shows domain classes and their description. The empty ontology in Protégé contains one class called *thing*. The class *Thing* is the super class that represents the set containing all classes. All other classes are subclass of Thing [3]. New classes in ontology can be added using the *classes* of protégé s pressed

2.2 Properties

Properties are attributes of object in the ontology. Properties are also used to create a relationship between classes as shown in fig. 2. The properties widget of the OWL classes tab allows user's to view and create relationship between classes. It provides access to those properties that could be used by the instances of the current class. The properties are edited through the form shown in figure 4. This form provides a metadata area in the upper part, displaying the property's name, annotation, and so on, similar to the presentation in the class form.

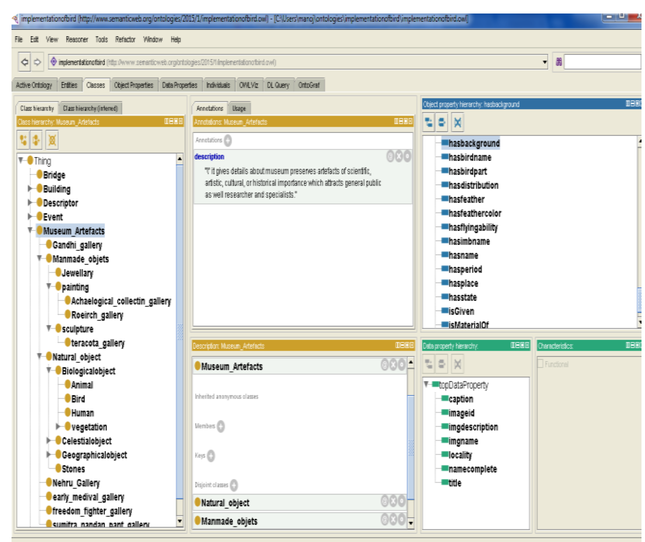

Figure 2: Classes and its properties

IJCSI International Journal of Computer Science Issues, Volume 13, Issue 4, July 2016
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org http://dx.doi.org/10.20943/01201604.4751

There are two main types of properties viz. Object properties and Datatype properties. Object properties link two instances together and present relationship whereas datatype property is used to assign values and link an instance to an XML schema datatype value or an RDF literal. OWL also has another property named annotation property, which is used to add metadata to classes, individuals and object/datatype properties [8].
Table 1 depicts some properties being used to define domain classes. The characteristics of properties are listed in table 2. Each attribute of a property is independent and has its own characteristics.

Table 1: Properties used in Museum Artefacts

S.No.	Name of the Properties
1	hasPeriod
2	hasName
3	PeriodOf
4	CreatedBy
5	PresentedTo
6	MadeOf
7	PartOf
8	ParticipatedIn
9	UsedAs
10	WrittenBy

Table 2: Characteristics of Properties

S.No.	Name of the Characteristics
1	Functional
2	Inverse Functional
3	Transitive
4	Symmetric
5	Antisymmetric
6	Reflexive
7	Irreflexive

The super class of domain ontology is museum artefacts class from which all other classes are derived. The ontology provides understanding of the artefacts by decomposing complex items into their constituent logical categories including independent substances, dependent items such as attribute and properties, temporal items such as events and processes, context sensitive perspectives and various form of relation like Internal relation, External relation, Grounded relation, Intentional Relation, Existential Relation[17]. Relations are defined to capture additional information regarding the person who took part in an event, and the corresponding location as well as the way the various objects and events are related to each other. The main class of the ontology is Museum artefatcs, which includes all other classes in the domain. It describes common properties like date and place of the artefacts. The sub-classes model variety of events that could be depicted in museum artefacts like political, social events, official, personal, historical, etc., as well as events associated to natural phenomena such as rain, snow, water, sunset, sunrise, etc. The museum artefacts relate to these subclasses through the property hasPeriod and subclasses relate to artefacts through the property isPeriodOf. The subclass Sculpture, Jewellary, Terracotta related to the Manmade class through the property hasMadeOf and the class Manmade is related to the class Sculpture, Jewellary, Terracotta through the property is isMadeOf. These two properties are inverse to each other. The classes Painting and Roerich gallery related through creates and isCreatedBy. This property is functional and irreflexive.

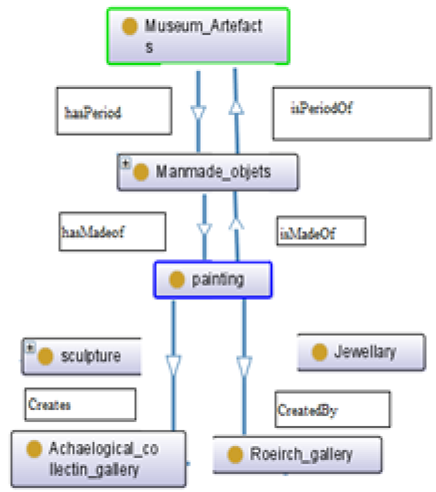

Figure 3: Property description using Classes

The properties can be edited using the properties tab selecting either object properties or datatype properties. Annotation can also be added to the properties in order to describe it. To create an object property switch to object properties tab and use the add object property button.
To set characteristics, the check box next to the property is checked. To specify properties protégé requires the specification of domain and range. In case of *creates* property shown in fig. 3 the domain is painting and its range is Roerich and the inverse of this will be isCreatedBy.

3 Open World Assumptions

OWL uses an open world assumption that if a statement cannot be proven true using current knowledge, we cannot draw the conclusion that the statement is false. The assumption denotes lack of knowledge. The consequence is that if two classes Man-made and Natural are not defined as disjoint then it can have the instances in common. The disjointness in classes plays a vital role in each of the class description. Creating a class and making it complement to another is done here. In other words, there is an assumption that we do not have all of the information and there may, or may not, exist information that makes something true.

3.1 Using Reasoner

The consistency can be checked through the reasoner. Protege supports many reasoners. This paper uses FaCT++ as its reasoner. FaCt++ is tableaux based reasoner for expressive Description Logics. It covers OWL and OWL2 description based ontology language. Reasoning means to infer new knowledge from the statements asserted by an ontology designer. The illegal mistakes committed by the developer are spotted out by the reasoner.The problem that is faced when the artefacts ontology is developed is due to wrong setting of property characteristics, the reasoner displays error messages like inconsistent ontologies.
Reasoning capabilities are exploited to detect logical inconsistencies within the ontology. The error may occur while setting characteristics. For example, if a property is described both as asymmetric and reflexive. The consistency check can help developer in manner while constructing the ontology.
The reasoner is not able to handle full expressivity. The OWL specification distinguishes between OWL full and OWL DL to indicate tractable language elements to reasoner. Ontologies which use metaclasses which is an OWL full element cannot be classified. The conversion of OWL full to OWL DL can be made using the classifier. Complete OWL full syntax is not supported by protégé.

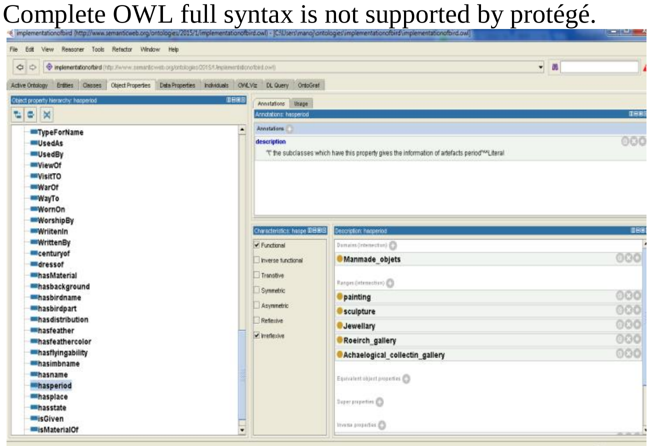

Figure 4: Object property

3.2 RDF/XML Rendering

The structure of any expression in RDF consists of triples, each consisting of a subject, a predicate and an object. A set of triples is called an RDF graph. This can be illustrated by using node and arc node link. In order to avoid conflict of conversation among different languages for describing ontology, there is a need of a common language specification. XML has been used for this purpose since it has standards for data exchange. OWL ontology is most commonly serialized using RDF/XML syntax. Figure 5 represents the RDF/XML rendering of ontology.

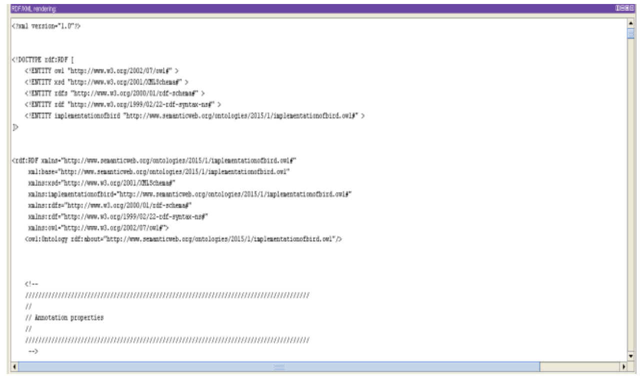

Figure 5: RDF/XML rendering

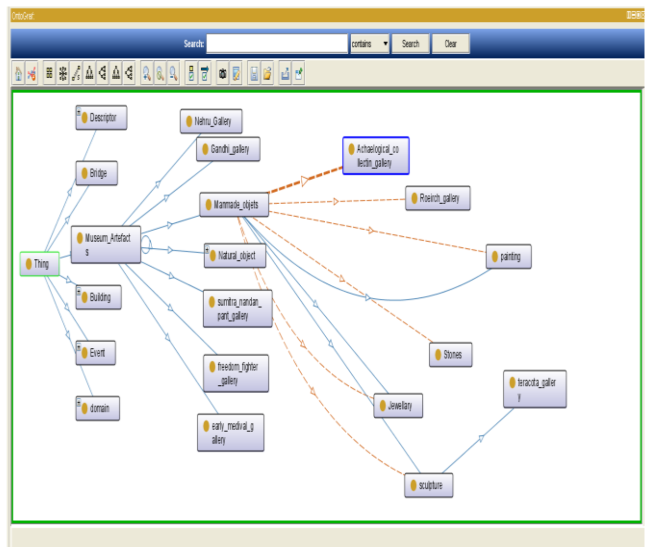

Figure 6: Museum Artefacts Ontology

4. Test Environment and Dataset

For the purpose of experimental evaluation, we have prepared a dataset with cooperation of Allahabad Museum. The dataset consists of 1200 images of artifacts from 11 galleries and ground truth annotation generated using the domain ontology. The graphical view shown in fig. 6 describes the super class and some of the subclasses of Museum Artefacts. The graph shows the relationship between super class and sub classes.
Recognition of an object in a scene can be easily handled by previous knowledge on the context or the interpretation of an image. The existing set of labeled images (trained Dataset) is searched in response to a query. The query may be in the form of an image, text-based or combination of both. For example, given the query ‘ashes of Mahatma Gandhi’, the system will process it to get related categories and will search in those categories. The ontology model together with image instance data can be used in finding

out relation between selected image and other images in the repositories. The image retrieval is done using SPARQL [18]. SPARQL is based on SQL and has the capabilities for querying visual graph patterns along with their conjunction and disjunction. The following question can answered by our ontological system.

4. Conclusion and Future Work

We have created ontology for Museum Artefacts which is used to annotate images in museum. Protégé interface provides RDF/XML code by utilizing the annotated image and relationship. Ontology enriched image metadata will help in semantic retrieval of images. The retrieval process is discussed with the help of examples.

Acknowledgments

We thank Director, Allahabad Museum for permitting us to collect images to create a dataset. The authors gratefully acknowledges the financial support from Dept. of Science and Technology, Govt. of India under Grant No SR/FTP/ETAY36/2011
.

References

[1] H. Knoblauch, O. Dameron, M. A. Musen,"Weaving the Biomedical semantic web with the protégé OWL plugin", Stanford Medical Informatics, Stanford University, Stanford, CA.

[2] A.Smeulders, M.Worring, S. Santini, A. Gupta, and R. Jain, "Content based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal.Mach.Intell.* vol.22, no.12, pp.1349-1380, Dec.2000.

[3] M. Horridge,"A Practical Guide to building OWL Ontologies Using Protégé 4 and C-ODE Tools "Edition 1.2.

[4] Y. Rui, T.S. Huang, S-F.Chang, Image retrieval: current techniques, promising directions, and open issues, J. Visual Commun. Image Representation, 1999, pp.39-62.

[5] N. Ruan, N. Huang, W.Hong,"Semantic –Based Image Retrieval in Remote Sensing Archive: An Ontology Approach", Geosciences and Remote Sensing Symposium, 2006.IGARSS 2006, pp.2903-2906.

[6] M. Dean and G. Schreiber, "OWL Web Ontology Language Reference", W3C Recommendation, 2004.

[7] The Protégé Ontology Editor and Knowledge Acquisition System, Available protégé.stanford.edu.

[8] M. C. Daconta, et al., The Semantic Web: A Guide to the future of XML, Web Services, and Knowledge Management, Indianapolis, IN: Wiley Publishing, Inc., 2003.

[9] N. Lammari, E. Metais, Building and maintaining ontologies: a set of algorithm, Data and Knowledge Engineering 48,155-176, 2004.

[10] P.A.S. Sinclair, S. Goodall, P.H. Lewis, K. Martinez, M. J. Addis, Concept browsing for multimedia retrieval in the SCULPTEUR project, Multimedia and the Semantic Web Workshop. *Annual European Semantic Web Conference (ESWC)* 2005:28-36.

[11] D. D. Dhoble, B.S. Patil, V. R. Ghorpade,"Semantic understanding of Image content", International Journal of Computer Science Issues, vol.8, Issue 3, No.2, May 2011.

[12] L. Hollink, A.T. Schreiber, J. Wielemaker, B. Wielinga, Semantic Annotations of Image Collections. *Workshop on Knowledge Capture and Semantic Annotation (KCAP)*, Florida, 2003.

[13] J.R. Smith, S.F. Chang, VisualSeek: a fully automatic content based query system. *Proceeding of the fourth ACM International Conference on Multimedia*1996:87-98.

[14] K.W. Park, J.W. Jeong, D.H. Lee, Olybia. *Ontology-Based Automatic Image Annotation System Using Semantic Inference Rules*. In R. Kotagiri et al. (Eds.) LNCS 4443, DASFAA; 2007. p. 485–496.

[15] T.J. Siddiqui, U.S. Tiwary, Words and Pictures: An HCI Perspective. *Proceeding of the First International Conference on Intelligent Human Computer Interaction* 2009; 59-70.

[16] C. Lee, Y. Kao, Y. Kuo, M. Wang, Automated ontology construction for unstructured text documents. Data & Knowledge Engineering 60 pp.547- 566, 2007.

[17] M.Giebelhausen, The architecture of the museum: Symbolic structures, urban context, Manchester University Press, Manchester (2003)

[18] K.Tzortzi, Museum Space: Where architecture meets museology.

[19] A.T. Schreiber, B. Dubbeldam, J.Wielemaker, B.J. Wielinga. Ontology-Based Photo Annotation. *IEEE Intelligent Systems*2001; 16:66-74.

[20] F.A. Mahdi,A. Ibadi,MIRS:Museum Image Retrieval System using Most Appropriate Low-Level Feature Descriptors, International Journal of Computer Science Issues, vol.11, Issue 5, No.2, September 2014 1694-0784.

Manoj Kumar Sharma is Ph.D student in department of Electronics and Communication, University of Allahabad, Allahabad, India. His main research interests are in the area of Image retrieval system, Computer Vision, his current research is image retrieval and content based retrieval of artefacts.

Dr. Tanveer J Siddiqui is Associate Professor at Allahabad University. Her current research area is language technology of information retrieval.