Using ontology for building distributed digital libraries with multimedia contents

HIRANMAY GHOSH, Tata Consultancy Services, Plot No. 249 D&E, Udyog Vihar Phase IV, Gurgaon – 122 016, India
GAURAV HARIT, Department of Computer Science and Engineering, Indian Institute of Technology, Kharagpur – 721 302, India
SANTANU CHAUDHURY, Department of Electrical Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi – 110 016, India

World Digital Libraries 1(2): 83–100

Abstract

This paper presents a new approach to build distributed digital libraries with multimedia contents. The authors propose a new scheme for media feature based concept modelling to address the limitation of traditional ontology based multimedia retrieval systems. The perceptual models can be used for semantic query processing using standard MPEG-7 media content descriptions. The authors have defined a new ontology language M-OWL (multimedia web ontology language) to support this perceptual modelling. M-OWL is an extension to the OWL (web ontology language) with new constructs for formal representation of the media properties of the domain concepts. It supports probabilistic evidential reasoning for robust concept recognition in multimedia documents. The separation of perceptual modelling of concepts from the repository architecture enables seamless integration of diverse multimedia contents. SOA (Service Oriented Architecture) is used to integrate large number of distributed information sources, each of which is modelled as an intelligent information agent. The authors have demonstrated the capability of the architecture by building a few research prototypes, namely a virtual encyclopaedia of Indian culture, a document image repository and a multimedia portal.

Email hiranmay.ghosh@tcs.com; gharit@cse.iitkgp.ernet.in; santanuc@ee.iitd.ac.in

Introduction

With the advent of multimedia technology, digital libraries are undergoing a silent revolution. Recent years have witnessed an exponential growth in different types of media contents, such as photos, music and videos, in the digital libraries. The primary reason of their popularity is the enhanced user experience and the deeper impacts on the human minds that they create. Moreover, digital archives for legacy documents often use document images. Unavailability of OCR software for many vernacular languages and the need to preserve the look and feel on the ancient manuscripts are the prime motivation for archiving the legacy documents in their image form.

The recent developments in information and communication technologies provide affordable acquisition, processing, storage and dissemination of multimedia contents. However, with the growth of digital multimedia contents, users face a great challenge in coping up with the large volume and the complexity and diversity of the media forms. The digital media libraries generally provide a rudimentary metadata and annotation based retrieval capability. However, metadata and textual annotations cannot capture the complete semantics of media forms. Selection of intended documents generally requires visual and aural inspection, resulting in overwhelming cognitive load on the researcher, especially for continuous media such as speech, music and video. Availability of large on-line storage and high-speed processors presents new opportunities of enhanced user experience through on-line content-based interaction.

Content-based access to digital media unfolds a whole lot of new research problems. User interaction with the media contents at a semantic level requires contextual interpretation of media data. Several research groups have reported success in content-based retrieval of image and video data in specific application domains, such as medicine, sports and news (Chu, Hsu, Cardenas, et al. 1998, Naphade and Huang 2001, Assfalg, Bertini, Colombo, et al. 2002, Chen 2004). These systems use multi-tier domain knowledge to interpret the media features. The domain models used in these systems are usually tailored to the repository contents (media forms) and architecture. A multimedia application often needs to integrate several media collections. For example, an encyclopaedia on Indian cultural heritage can be built over the contents of several collections on the Internet, collated by different research and commercial groups. These collections are available in a variety of media forms (typically, still image, video and music) and hosted over different technology platforms. The existing knowledge-based multimedia information systems cannot scale up to accommodate such dissimilar information sources.

Several research groups (Hunter 2003, Hammiche, Benbernou, Hacid, et al. 2004, Petridis, Bloehdorn, Saathoff, et al. 2005, Tsinaraki, Polydoros, Kazasis, et al. 2005) have proposed use of ontology in semantic interpretation of multimedia data in collaborative multimedia and digital library projects. While ontology is a useful tool for modelling a conceptual domain, it has not been designed to model multimedia data that is perceptual in nature. In the current approaches, specific computer vision algorithms are used to recognize pre-defined objects or events of interest in the media documents to generate automatic annotations. Domain ontology is used to interpret these conceptual annotations in specific query contexts. As a result, contextual information, which is present in the media form but are not a-priori recognized through computer vision techniques, gets ignored. Moreover, it is very difficult to articulate the desired aesthetic properties of a media instance, such as painting, music and movies, during query formulation. True semantic processing of multimedia data requires in-context interpretation of the information encoded in the media instance itself.

Our research with multimedia contents addresses these issues. The authors have proposed a framework for content-based interaction that enables seamless interoperability of different media collections at semantic level. The crux of our approach is to bridge the semantic gap between the conceptual and perceptual worlds using ontology, which is independent of collection architecture. They view the lack of support for perceptual modelling to be a severe limitation of the existing ontology languages in context of semantic processing of multimedia data. The authors propose that the ontology for multimedia applications should enable modelling of the concepts in terms of their perceptual manifestations in different media forms. This perceptual model so derived can be used as a tool for concept interpretation in multimedia artefacts. They recognize that a concept can have different alternative media manifestations, which lead to uncertainties with perceptual modelling of concepts. Thus, ontology designed for multimedia applications should support reasoning with uncertainties, which is not available with Description Logic based crisp reasoning schemes around contemporary ontology representation schemes, for example, OWL. Moreover, the authors note that the media property descriptions have some special semantics. For example, a monument built with a certain building material in likely to 'inherit' the latter's colour and texture properties. A multimedia ontology should enable reasoning with such semantics of media property descriptions.

In this paper, the authors present a novel way of using ontology to represent the relationship between the conceptual and the perceptual worlds. They have proposed M-OWL (Multimedia Web Ontology Language) as a language for formal definition of multimedia ontology. The media events and conceptual entities are described using contextual, structural (spatial and temporal relations) and perceptual models in M-OWL. M-OWL is a standards-based approach. Syntactically, it is an extension of the OWL (Web Ontology Language). The media features are described using extended MPEG-7 constructs. However, the authors have proposed a new Bayesian (evidential) reasoning scheme with M-OWL for concept recognition. Evidential reasoning systems can accommodate the inherent uncertainties in the media world and can produce more robust results than the deductive reasoning systems.

The separation of multimedia ontology from the document collection architecture offers the flexibility to integrate heterogeneous information sources and develop scalable systems. The perceptual specification of a concept comprises a redundant set of media pattern specifications and observation of a non-redundant subset results in concept recognition. The flexibility in choice of concept recognition strategies provides the key to integration of distributed multimedia collections. The authors propose a SOA (Service Oriented Architecture) involving several independent service elements as a convenient tool for building flexible and scalable systems. Each service element in the system has been modelled as an intelligent information agent and performs knowledge-based multimedia information processing tasks. The framework enables realization of a robust distributed digital library system, where several autonomous agents encapsulating information, media processing and ontology resources can coexist and interact.

The authors have implemented a few applications around this architecture. HeritAge is a virtual digital library. This is an Internet based solution to provide semantic access to a number of websites hosting information about traditional Indian arts and culture in different media forms. The open agent based architecture makes it readily extensible to incorporate additional domain knowledge as well as additional document collections. Heritage+ extends this framework to include legacy document image collections. Generic models of the document images help in classifying the documents and extracting specific sections of interest in this system. Multimedia Explorer is an immersive multimedia contents exploration environment, where video segments are hyperlinked based on the basis of conceptual similarity at multiple levels of abstraction to facilitate contextual navigation.

The rest of the paper is organized as follows. Section 2 provides an overview of digital multimedia library initiatives around the world. Section 3 explains the requirements for multimedia ontology language and introduces M-OWL in this context. This section explains the evidential reasoning adopted with M-OWL for concept recognition too. Section 4 explains the use of M-OWL in providing an integrated retrieval environment for different multimedia data forms in a digital library. Section 5 provides an overview of the web service based architecture and software agents for integrating distributed and heterogeneous repositories. Section 6 provides some application examples of this framework. Section 7 concludes the paper.

Digital multimedia libraries around the world

Today, there are several initiatives to create digital libraries with multimedia contents in different domains, such as cultural heritage, arts and crafts and news services. An exhaustive review of all such initiatives is beyond the scope of this paper. The authors provide a few examples of diverse initiatives around the globe.

The Imperial College of London together with University of Waikato, New Zealand and other institutes has taken up a project to put together digital contents from the archives of the BBC, the British Library, the New Zealand Digital Library and the Victoria and Albert Museum (Bainbridge, Browne, Cairns, et al. 2005). DELOS (2008) is the European Commission vision for networked virtual libraries to enable anyone from their home, school or office to access the knowledge contained in the digital collections created by traditional libraries, museums, archives, universities, governmental agencies, specialized organizations, and individuals around the world. Global Memory Net (Chen 2004), supported by the NSF/ IDLP (International Digital Library Program), is conceived to be a model global digital library of cultural, historical, and heritage image collections. Internet Archive (2008) is building a digital library of video, music, audio and text and provides free access to researchers, historians, scholars, and the general public. Scottish Distributed Digital Library (2008) focuses on Scottish themes and contains multimedia artefacts like document images, digital photographs, scanned paintings and is maintained by a range of information organizations, such as archives, libraries, museums and private collectors.

Back in India, Kalasampada (IGNCA 2008) is a digital library of Indian cultural heritage containing digital images, video recordings and scholarly annotations. IGNCA (Indira Gandhi National Center of Arts) has created this collection in collaboration with National Informatics Centre. Ministry of Human Resource Development has set-up an Indian National Digital Library in Engineering Sciences and Technology Consortium, comprising reputed academic and research institutions in the country. The distributed digital library (Digital Library 2008) spans over the local collections of the participating institutes and provides links to other media-rich websites including national and regional newspapers, heritage manuscripts and information gateways on specialized subjects. (Das 2008) provides a comprehensive survey of digital library initiatives in South Asia.

Ontology for multimedia applications

For quite some time, information professionals have been using several tools, for example, thesauri, subject taxonomies, established classification systems and specialized classification schedules, for knowledge classification. Ontology is a further generalization of these tools. It is a formal representation of conceptualization of a domain in terms of concepts, their properties and their interrelations. (Carlyle 2002) provides a critical review of application of ontology for knowledge organization in libraries. With the proliferation of the Internet, W3C forum has undertaken the initiative of standardizing the ontology representation for websitebased applications. The OWL, standardized in 2004 after maturing through XML(S), RDF(S) and DAML+OIL is a result of that effort. Ontology in OWL (and some of its predecessor languages) has been successfully used in establishing semantics of text in specific application contexts.

Semantic processing of media data calls for ontology primitives that permit perceptual modelling of domain concepts. Ontology for multimedia applications should enable translation of the conceptual model of an object or an event into a media based model (and vice-versa) in an application or repository independent way. OWL and the other existing ontology languages do not support such modelling. In spite of their fundamental differences, the conceptual and the perceptual worlds are closely coupled with cause-effect relationships. For example, the concept ‘steam locomotive’ causes some specific media patterns, for example, its body shape, its characteristic whistle, and so on, to manifest in multimedia artefacts. Observation[1] of these patterns in a media artefact leads to the belief in the presence of the concept. The authors have proposed OM (Observation Model) as a representation of a concept in terms of such expected media properties. Figure 1 depicts a possible observation model for a ‘steam locomotive’. A concept can have many alternative manifestations in a media form, for example, many different possible body shapes of a steam locomotive. Thus, there is a good deal of uncertainty in the correspondence of concept and their manifestation features. An Observation Model encodes the uncertainty in the form of cause-effect probabilities and uses a Bayesian Network as the probabilistic reasoning mechanism to recognize concepts. Observation of a media pattern in an OM provides an elementary evidence for the

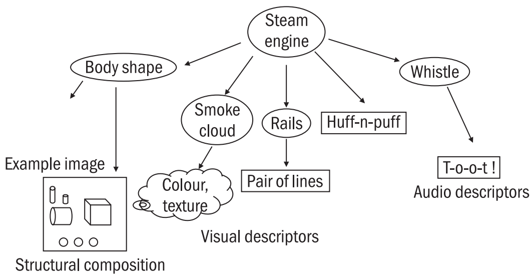


Figure 1 Observation model for a steam locomotive

[1] This word ‘observation’ is used in a more general sense of perception, to include media patterns other than the visual ones.

concept and causes belief revision in the Bayesian Network. A concept is recognized by the strength of accumulated evidence because of several observations.

Formulation of Observation Models that can be used concept recognition in multimedia documents requires some domain knowledge. The authors have defined a new M-OWL for encoding media properties of concepts and reasoning with them to facilitate construction of observation models of the concepts in a closed domain (Ghosh, Chaudhury, Kashyap, et al. 2007). M-OWL has been defined as an extension of the OWL to protect existing investments in ontology. M-OWL incorporates additional constructs for describing media properties of concepts. These constructs have been motivated by MPEG-7, which provides a flexible mechanism for describing media contents at multiple levels of abstraction. M-OWL extends MPEG-7 constructs for formal specification of spatial and temporal relations between media objects and events (Wattamwar and Ghosh 2008a). Moreover, M-OWL supports constructs for encoding uncertain causal relations between concepts and their media properties. While the authors used the approach followed in BayesOWL (Ding and Peng 2004) in encoding Bayesian prior and conditional probabilities, they have developed distinct reasoning mechanism for hierarchical relations, media property propagation and spatio-temporal relations that constitute multimedia ontology.

A digital multimedia library can make use of an M-OWL sports ontology for semantic interpretation of sports media archives. The authors illustrate encoding of a typical sports event, goalScore event in a football (soccer) match in Figure 2. The event has been defined

```
<mowl:STOp rdf:ID="followedBy">
    <hasCt><w>-INF 0 0 1 0.5 1 0 INF 0</w></hasCt>
</mowl:STOp>
<mowl:STOp rdf:ID="inside">
        <hasXp><v>-IN 0 0 0 0.5 0.5 1 1 3 1 3.5 0.5 4 0 INF 0</v></hasXp>
        <hasYp><v>-IN 0 0 0 0.5 0.5 1 1 3 1 3.5 0.5 4 0 INF 0</v></hasYp>
</mowl:STOp>
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
<mowl:MediaFeature rdf:ID="ballShape">
        <Mpeg7…> A shape descriptor for ball </Mpeg7…>
</mowl:MediaFeature>
<mowl:MediaFeature rdf:ID="ballColor">
        <Mpeg7…> A color descriptor for ball </Mpeg7…>
</mowl:MediaFeature>
<mowl:MediaFeature rdf:ID="goalboxShape">
        <Mpeg7…> A shape descriptor for goalbox </Mpeg7…>
</mowl:MediaFeature>
<mowl:MediaFeature rdf:ID="cheerSound">
        <Mpeg7…> An audio descriptor for cheer </Mpeg7…>
</mowl:MediaFeature>
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
<owl:Class rdf:ID="ball">
        <hasMediaFeature>ballShape</hasMediaFeature>
        <hasMediaFeature>ballColor</hasMediaFeature>
</owl:Class>
<owl:Class rdf:ID="goalbox">
        <hasMediaFeature>goalboxShape</hasMediaFeature>
</owl:Class>
<owl:Class rdf:ID="cheer">
        <hasMediaFeature>cheerSound</hasMediaFeature>
</owl:Class>
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
<owl:Class rdf:Id="goalScore">
        <mowl:followedBy rdf:parseType="Collection">
                <owl:Class rdf:about="cheer" />
                <mowl:inside rdf:parseType="Collection">
                        <owl:Class rdf:about="goalbox" />
                        <owl:Class rdf:about="ball" />
                </mowl>
        </mowl>
</owl:Class>
```

Figure 2 M-OWL description for a goalScore event in sports ontology as a spatio-temporal composition of elementary media concepts as ((ball inside goalbox) followedBy cheer), where the elementary concepts ball, goalbox and cheer are characterized by their visual and audio properties and the relations inside and followedBy assume formal spatial and temporal semantics. There are four distinct sections in this description (separated by dashed lines in the diagram). The first one defines the semantics of followedBy and inside relations. The tag 'mowl:STOp' signifies the Spatio-Temporal Operator. It specifies the fuzzy membership function as a pair of mappings indicating respectively the input feature value and its corresponding fuzzy membership value. The second section defines the visual and audio features for identifying elementary objects, for example, the 'ball', the 'goalbox' and the 'cheer' in a video scene in the form of a template. The third section associates these features to the media objects. The fourth section defines the goalScore event in terms of the constituent media objects and the spatio-temporal relations. The ontological description translates to an Observation Model shown in Figure 3. The Bayesian probabilities are not specified in this example and some default values are assumed.

Moreover, an OM for a concept comprises not only the media properties of the concept but also the media properties of related concepts, for example, the presence of pair of lines signifying railway tracks provides evidence towards possibility of a steam locomotive. While M-OWL is syntactically an extension of OWL, it defines new semantics for the media property descriptions to cope up with such media property propagation. As further examples, an image instance depicting a monument built with marble is likely to exhibit the colour and texture properties of marble. On the other hand, since the 'Tajmahal' is an instance of a tomb, any example image of the 'Tajmahal' is also an example image for a tomb, since tomb depicts a broader class of monuments of which the 'Tajmahal' is a member. Figure 4 illustrates some 'downward' and 'upward' property propagation in the knowledge graph represented by an M-OWL domain description. These property propagation rules are quite distinct from property inheritance in frame-based knowledge representation and are necessary for construction of Observation Models. M-OWL supports constructs for assigning media propagation attributes to domain specific relations to enable such

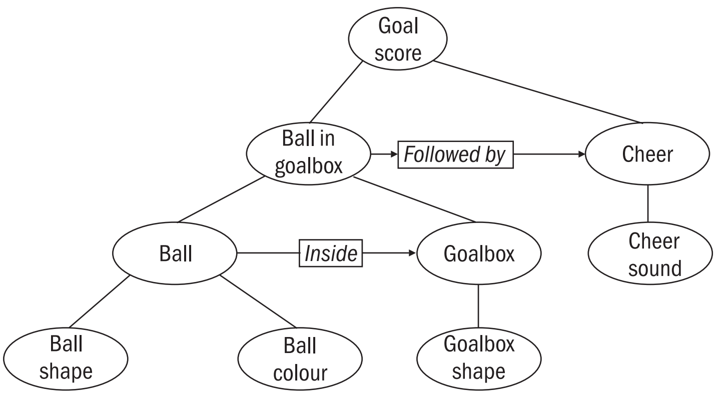


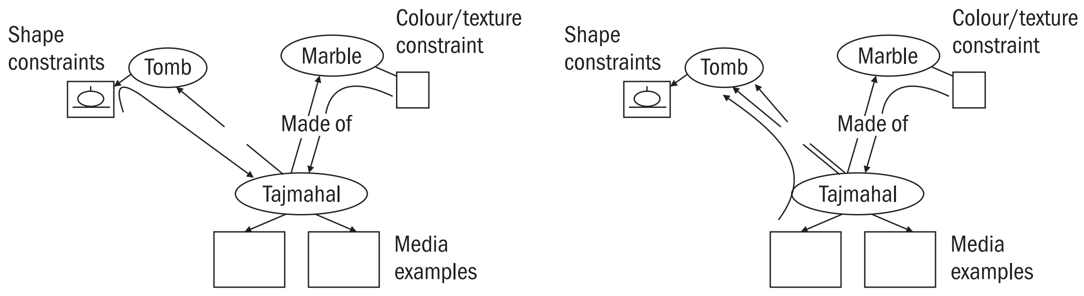
Figure 3 Observation Model for goalScore event in football (soccer)

reasoning. To summarize, the authors state that M-OWL differs from OWL in terms of the Bayesian reasoning framework and the property propagation semantics.

M-OWL for multimedia data integration

A digital library can contain different media forms including music, video, speech, photographs and document images. In contrast with the existing digital library architectures, which segregate the different media forms and uses different methods to process them, an advantage of M-OWL is that it can be used to process the different media forms in a unified framework.

An important feature of an Observation Model derived from M-OWL ontology is its generic nature for perceptual description of a conceptual object or event. A concept can be described in M-OWL with its several different expected manifestations in various media forms. In general, an observation model includes all these different media properties. However, it is not necessary to observe all such media patterns to recognize a concept. When the belief in a concept is sufficiently reinforced with a few observations of the expected media properties, further observations may not add significantly to the belief value. For example, observation of a smoke cloud (visual) and a "t-o-o-t" (audio) can generate sufficient belief for a steam locomotive. Later observation of its body-shape does not add significantly to this belief. The authors refer to this phenomenon as the diminishing value for the evidences. Thus, there is a degree of redundancy in the observation model, which leads to the flexibility in choosing an appropriate set of expected media patterns in context of specific media form. The designer of the observation model has to choose an appropriate set of patterns keeping in consideration the evidential strength of the media patterns towards the concept as well as the feasibility and computing overheads of the pattern recognition task. For example, a retrieval system need to select some visual features of a steam engine in order to recognize one in a collection of still photographs, while it may substitute some of these visual features with a few prominent audio features while dealing with a video collection. (Ghosh 2004) provides an algorithm for selecting an appropriate set of media patters for observation in context of a given retrieval task and repository architecture. The monetary



(a) Downward flow of media properties (b) Upward flow of media examples

Figure 4 Media property propagation in multimedia ontology cost for third party services in a service oriented architecture can also be a consideration for selecting appropriate media patterns for retrieval.

Many of the existing multimedia collections extract some media features and generate a content description in order to facilitate access. MPEG-7 (Manjunath, Salembier, and Sikora 2003) has been standardized as a language for content description of multimedia documents. An M-OWL based information system can interpret such descriptions, wherever they are available.

Software agents for distributed libraries

The success of the Web can largely be attributed to the massive decentralization of information sources. There is a plethora of technologies for representing, organizing and accessing the information on the Internet. Every author on the web has absolute freedom to choose and organize the contents and every web publisher has the freedom to choose a technology with minimum compatibility requirements. This has resulted in huge volume of information being freely and quickly published on the Internet. This decentralized knowledge model applies to organized digital libraries as well. Because of content ownership and administrative issues, it is generally convenient for individual research groups to build their own digital collections independently, rather than to collaborate with each other to build a central knowledge repository. The authors found this approach being adopted by the partners of the Indian consortium of digital libraries as well as in the European DELOS project. While there are significant thematic overlaps across the collections, the owners of the individual digital libraries are generally free to use any technology, to choose any repository organizations and different retrieval schemes, some of which can be knowledge-based. Thus, it is pragmatic to assume that digital collections will have a fragmented existence over dissimilar and distributed repositories. However a researcher or a user would always like to have an integrated view of these distributed collections. It is therefore necessary that that the systems design should offer a seamless integration of a large number of independent collections, despite their heterogeneity.

Agent based systems provide a suitable framework to design and develop such massively distributed knowledge based systems. Figure 5 shows a generic agent based architecture for digital media libraries. In this architecture, each of the individual document collections are treated as an independent knowledge-based information service, designated as Repository Agent in the figure. Semantic integration of these individual services needs some additional system components. User Interface Agents cope up with the HCI (Human-Computer-Interface) issues. They employ different mechanisms to interact with the users to suit user's cultural and educational background, expertise level, physical challenges and other contextual needs. Ontology Agents interpret abstract user queries in terms of observational requirements. The interaction between the knowledge elements encoded in the Ontology Agents and the Repository Agents poses an interesting research issue. A Coordination Agent handles the task of the query processing. It engages and interacts with a subset of available Ontology and Repository Agents to realize the retrieval services. The selection of these agents is based on the query and user context. There is also a provision for the Repository Agents to engage third party Media Agents for specific intelligent media processing tasks, such as face recognition, either during indexing or query processing phase. Directory Agents permit dynamic discovery of agents with specific capabilities.

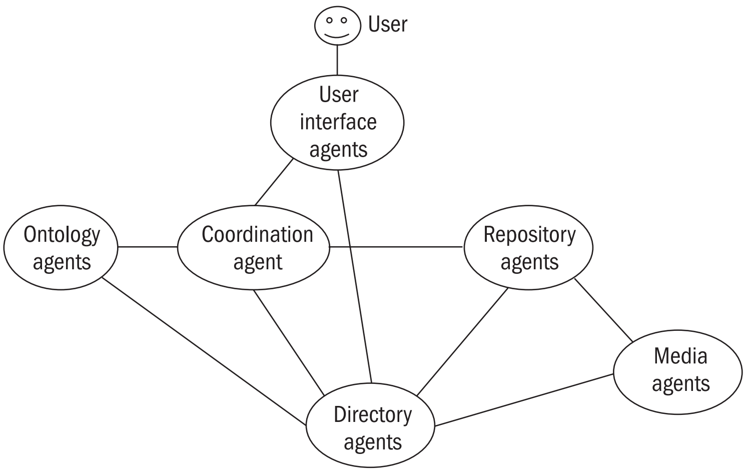


Figure 5 Agent based architecture for distributed digital library

These agents are discussed in detail later in this section.

The main advantage of employing agent based system architecture is the flexibility of design. An agent-based system provides very loose system architecture, so that it is possible to design each service component independently and autonomously. Moreover, agent based systems permit dynamic addition and deletion of agents without disruption of services. The existing agents automatically regroup on such occasions to provide uninterrupted system functionality. This feature provides robustness to the distributed digital library architecture with many independent library services, some of which may be withdrawn or reorganized with time and new services may be added without a global knowledge. The system sustains itself through a mutual message exchange, rather than being controlled in a centralized fashion.

Since the different agents in the system are independently developed and deployed on a variety of hardware platform and software framework, they need a common framework for communication. WSA (Web Services Architecture) (W3C 2004) provides a standard means of interoperating between such heterogeneous agents. In this architecture, each agent is modelled to have realized a set of 'services', which can be accessed by the other agents by exchanging messages following a definite set of protocols. Simple Object Oriented Protocol (SOAP) defined as a part of this architecture standardizes the method of exchanging XML messages between two agents. However, WSA does not specify the interpretation of the XML message contents and an agent's action on receiving a message. The Foundation for Intelligent Physical Agents (FIPA)[2] defines a set of protocols that guide such interactions between the agents. The authors have used some of these protocols to realize the various digital library services. Typically, the services include ontology based query expansion (formulating the observation model), formulating the retrieval strategy for a document collection, executing the retrieval plan, and so on. These high-level services are often decomposed to smaller service components, such as matching visual query patterns.

In a decentralized digital library environment, new collections can be deployed and some of the existing collections can be reorganized anytime. Same considerations holds good for the different ontology and media analysis services. Provision for dynamic agent team formation required an intelligent service discovery mechanism. The authors have created a set of Directory Agents in the system to facilitate this function. These agents provide registration and capability based search services. Whenever a new agent is commissioned in the system, it registers its services and the method to invoke them with a Directory Agent. Web Services Description Language (WSDL) (W3C 2001) is the standard XML format used for the purpose. Whenever an agent requires the services of other agents, it utilizes the search service of a Directory Agent by specifying the required capability. The Directory Agents do an intelligent matching of the required services and the capabilities of the registered agents and propose a set of relevant agents which can cater to the requirement. The Directory Agents have a federated architecture. An agent can access the registration or search services of any of the Directory Agents of the system. In general, the Directory Agents consult each other to satisfy a search request.

The overall retrieval model in this agent based distributed digital library architecture is depicted in Figure 6. The Observation Model of a concept created by the ontology services is refined to optimal observation plans for concept recognition for the individual document collections using a distributed planning process (Ghosh and Chaudhury 2004). An optimal observation plan for a repository comprises a set of observations that can provide sufficient confidence at the lowest possible computational cost. Different collections have different optimal observation plans, since the feasibility and the execution cost of making an observation depends on the repository contents and data-models. The observation plans are executed at the collections to yield a set of relevant documents as results. The results from individual collections are combined to form the final result set.

[2] http://www.fipa.org

Application examples

The authors have developed a number of applications involving different media forms, for example, still images, video, document images, and the like, with this perceptual modelling technique using M-OWL. They provide a brief description of some of these applications in the following text.

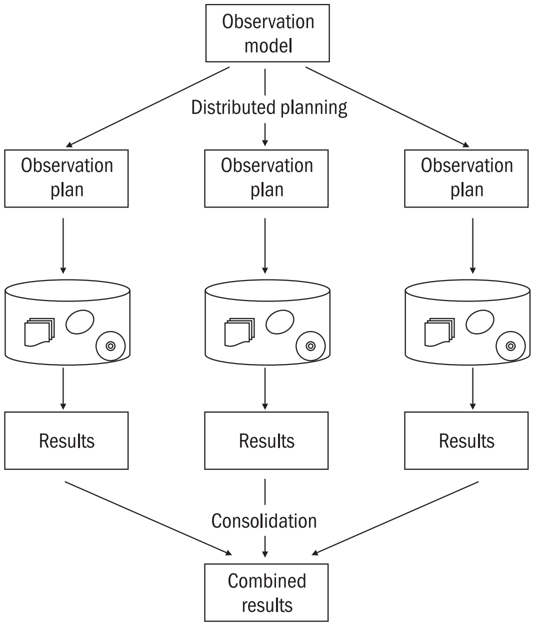


Figure 6 Retrieval in distributed digital library

HeritAge

HeritAge is a virtual encyclopaedia of Indian cultural heritage. There are quite a few websites that contain exhibits of Indian cultural heritage. Some of them, for example, Kalasampada from IGNCA, represent research, while some others, for example, Shalinarts.com and Khazana.com represent commercial interest. These websites have different organizations and contain different media forms. For example, Kalasampada is a collection of digital videos depicting Indian classical music and dance, ShalinArt.com contains digital images of Indian folk-paintings on sale, TheHorizons.com contains descriptions with embedded digital images of arts and crafts as well as classical music and dance, and so on. Though all of the sites can be publicly accessed, locating specific information about Indian Art and Culture from these sites with available Internet search tools for text, image and video requires lots of perseverance because of their heterogeneity. Moreover, descriptive queries such as ‘folk-art from eastern India’ produce disastrous results, since the search engines have no clues to the characteristics of such items.

Effective interaction with such diverse collections is possible with Observation Models for content-based classification of artefacts in multiple media forms. The challenge here is the conceptualization of the domain of Indian arts and culture for creation of these Observation Models. HeritAge has been developed as a proof of concept to establish the feasibility for such application. The authors have encapsulated a few representative Internet based collections to integrate in a virtual encyclopaedia environment. The authors have developed a domain model for a small section of Indian classical arts and historical monuments with several media properties and examples. This domain description does not exclude text, which is also an important media form in many of these collections.

Figure 7(a) depicts the query interface of HeritAge. A user can specify the intended documents in different ways, for example, with a set of pull-down menus, with free text expressions, and URL’s of example media objects (still images). The User Agent encodes the query in a generic query language and interprets the query to build an observation model. The Observation Model is generated by the Ontology Agent and is then used to generate Observation Plans for the different collections by making use of a distributed planning algorithm. Described in section 4 A set of Media Agents have been deployed for colour and texture based image and video classification. Figure 7(b) depicts search results for a query, paraphrased as ‘Carnatic vocal music’. The system retrieves a set of web pages, still images and videos, depicting Carnatic vocal recitals and information about the eminent artists from the different encapsulated websites. Figure 7(c) depicts the search results for ‘Kali in eastern Indian folk painting’. The query has been specified with textual description and an example image. The system produces a number of painting examples from several commercial and research websites, many of which were found to be relevant. This prototype demonstrates integration of repositories containing multiple media forms and efficacy of results achieved through collaborative evidential reasoning, despite use of simple low level feature recognition algorithms.

HeritAge+

The economic feasibility of maintaining large document image databases has created a tremendous demand for indexing and efficient access to document image collections. Many of the earlier document image management systems (Gatos, Mantzaris, Perantonis, et al. 2000, Altamura, Esposito, and Malerba 2001,

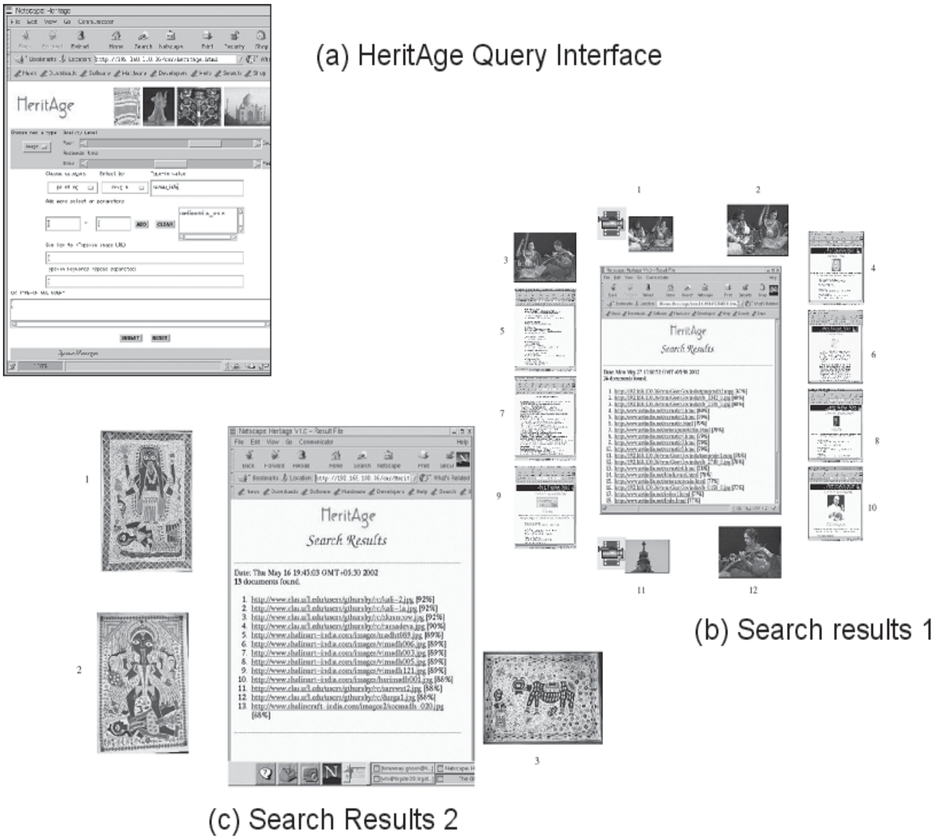


Figure 7 HeritAge user interface and search results

Appiani, Cesarini, Colla, *et al.* 2001, Wenzel and Maus 2001) transform document images to XML format by exploiting the document layout model or the logical structure, together with OCR techniques. Such techniques cannot be used for legacy documents in those Indian scripts for which reliable OCR technology is not available. The problem of word-image based document image indexing, that is, searching for keywords in document images using image properties, has received considerable attention in the past (Doerman 1998). However, conceptual access to document pages requires modelling the domain (pertaining to the document category) specific concepts in terms of the document layout, features extracted from picture components, and word image characteristics. The domain knowledge can be used to process, analyze, identify, and label the logical components in a document image. In addition to the component properties, the authors also include relations between components as part of domain knowledge. This facilitates access to document pages as a set of logically related components.

Heritage+ extends HeritAge to include conceptual search and navigation in document images archives. A document image archive is a digital library of scanned documents of

Figure 8 Logical decomposition of document image: The Hindu Crossword puzzle

various types, for example, classical literature, magazines, newspapers, pictures, and other miscellaneous documents, in several Indian languages. M-OWL can represent a document image structure in terms of the spatial layout and the image characteristics of its components. The authors use M-OWL to encode the domain knowledge, which includes abstract concepts and relations, a set of keywords related to the concepts and their observable forms (GFG feature or envelop curve feature) in document images in a number of Indian languages. M-OWL is also used to model the document features, for example, the layout and compositional structure together with the observable image properties. Figure 8 depicts the logical decomposition of a typical newspaper page. It is also possible to model the words in native languages as concepts and define their media properties in terms of spatial and sequential relation of some shape primitives (Geometric Feature Graph) using M-OWL, as shown in Figure 9. Observation of the connected shape primitives with some confidence, is used in robust recognition of the words in the native language (Harit, Chaudhury, and Paranjpe 2005). The structural properties of the identified components in the document page can be used to classify document images into generic categories.

Query processing in Heritage+ involves identification of document class (for example, sport news) followed by concept spotting (for example, Tendulkar). Document images in a collection are segmented into pages, labelled with semantic descriptors using document ontology. These pages are hyperlinked based on the relations between the page descriptors, to facilitate navigation. Personal annotations by a user complement the global annotations in the context of user initiated search and navigation requests. Concept spotting in documents uses domain ontology to expand the query concept into expected media features of the alternative keywords that signify the concept. These media features are then searched in the document images to identify the keywords and hence results in

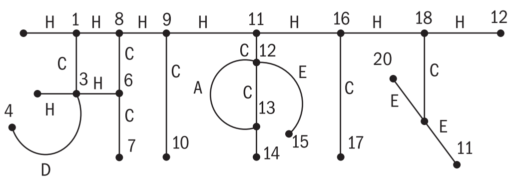


Figure 9 Modelling of words as a sequence of shape primitives

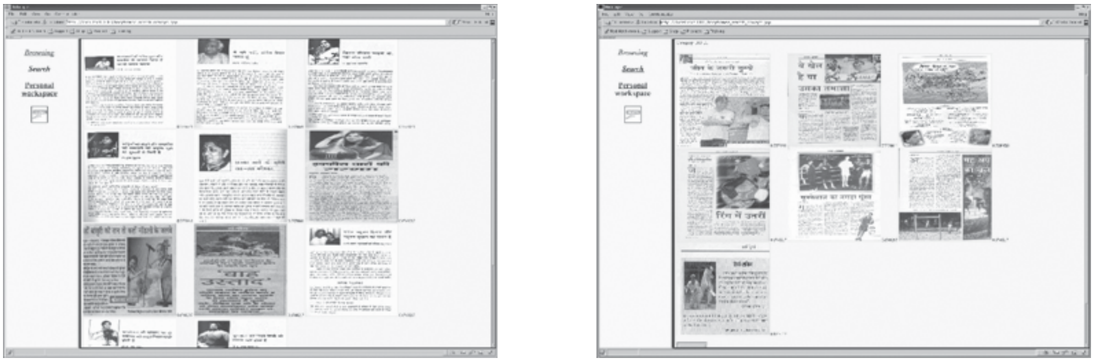


Figure 10 HeritAge+ document image archive

concept recognition. Figure 10 shows the query interface and the search results for 'Classical Music' with Heritage+.

Immersive Multimedia Exploration Environment

Multimedia Explorer (Wattamwar Mishra, and Ghosh 2008b) is a prototype immersive multimedia exploration environment. It integrates a user-friendly workbench for authoring content description of video documentaries and provides facility to navigate in a video library. It decomposes a video into constituent scenes and extracts their audio-visual properties to create a user-friendly content authoring environment, where the editor can create scene and object based annotations. Manual annotations are complemented with automated inferencing of the spatio-temporal interactions and conceptual relations with entities that manifest in the video. Much of the semantics associated with video data are contextual and characterized by uncertainty. For example, a low level concept like a 'slow moving subject' can be recognized if the location of the subject does not show much variation in time. The presence of one or a few slow moving subjects in a video can be interpreted as a dancer or a group of dancers (Harit and Chaudhury 2006a, Harit and Chaudhury 2006b) with a certain level of confidence in a video pertaining to Indian classical art. The resulting scene-level video description comprising annotations, audio-visual features, and space-time location of the media objects are stored in an MPEG-7 compliant database.

Multimedia Explorer utilizes this fine-grained video description together with intelligent query processing and ranking algorithms to facilitate a scene description based retrieval of video segments. It creates dynamic contextual presentations with video fragments (and their descriptions) using SMIL

Figure 11 An immersive multimedia exploration environment

(Synchronized Multimedia Integration Language). The Ontology also enables video scenes with semantic similarity to be hyperlinked for contextual navigation. Figure 11 provides a screen-shot of Multimedia Explorer with Indian classical arts in its collection. Though presently realized for a video collection, the prototype implementation can be readily extended to other forms of media artefacts, for example, still images and music by incorporating appropriate tools to process these media artefacts.

Conclusion

The heterogeneity of multimedia data and repository architecture pose significant challenges in creating distributed digital libraries containing media artefacts. The authors have presented a new ontology representation scheme for applications dealing with multimedia assets and methodologies for ontology assisted interaction in distributed media collections to address these challenges. They have developed and adopted a multimedia ontology representation language M-OWL which enables perceptual modelling of a domain over and above the conceptual modelling that is supported with conventional ontology representations. The generic perceptual model enables use of content-based retrieval techniques for retrieval in heterogeneous media forms. The authors have defined an evidential reasoning scheme around M-OWL to cope up with the uncertainties with multimedia data interpretation. This model of interaction with media libraries is versatile and can account for different media forms and repository architecture that proliferate the Internet and local collections. They have demonstrated this capability by realizing working prototypes with different types of contents, namely still photographs, document images and video artefacts. This is a step ahead of earlier content-based multimedia systems, which can generally process a single media form and cannot integrate heterogeneous repositories. Our research paves the way for creating large-scale distributed digital libraries, integrating several collections over the Internet. The use of Web Services Architecture and Intelligent Agent Technology leads to scalability and dynamic growth of the system with the advantage of decentralized control.

References

Altamura O, Esposito F, and Malerba D. 2001
Transforming paper documents into XML format with Wisdom++
International Journal of Document Analysis and Recognition (IJDAR), **4**: 2–17

Appiani W, Cesarini F, Colla A, Diligenti M, Gori M, Marinai S, Soda G. 2001
Automatic document classification and indexing in high-volume applications
International Journal of Document Analysis and Recognition (IJDAR), **4**: 69–73

Assfalg J, Bertini M, Colombo C, Del-Bimbo A. 2002
Semantic Annotation of Sports Videos
IEEE Multimedia Magazine, **9**(2), 52–60

Bainbridge D, Browne P, Cairns P, Rüger S, Xu Li-Qun. 2005
Managing the Growth of Multimedia Digital Content
ECRIM News Number 62. Details available at <http://www.ercim.org/publication/Ercim_News/enw62/EN62.pdf> last accessed on 5 September 2008

Carlyle A, 2002
Document Ontologies in Library and Information Science: An Introduction and Critical Analysis
Knowledge Technologies Conference 2002, Seattle, WA, USA

Chen C C. 2004
Global Memory Net: Potential and Challenges for Archiving and Sharing Cultural and Heritage Resources
Keynote address at *International Conference on Digital Libraries 2004: knowledge creation, preservation, access, and maagement.* New Delhi

Chu W W, Hsu C C, Cardenas A F, Taira R K, 1998
Knowledge based image retrieval with spatial and temporal constructs
IEEE Transactions on Knowledge and Data Engineering, **10**(6): 872–888

Das A K. 2008
Open Access to Knowledge and Information. Scholarly Literature and Digital Library Initiatives—The South Asian Scenario
UNESCO (United Nations Educational, Scientific and Cultural Organization)

DELOS. 2008
DELOS Network of Excellence on Digital Libraries
Details available at <http://www.delos.info/>, last accessed on 5 September2008

Digital Library. 2008
Digital Library of India
Details available at <http://www.digitallibrary.ernet.in/>, last accessed on 5 September 2008

Doermann D. 1998
The Indexing and Retrieval of Document Images: A survey,
Computer Vision and Image Understanding CVIU, **70**(3): 287–298

Gatos B, Mantzaris S, Perantonis S, Tsigris A. 2000
Automatic page analysis for the creation of a digital library from newspaper archives
International Journal on Digital Libraries, **3**: 77–84

Ghosh H and Chaudhury S. 2004
Distributed and Reactive Query Planning in R-MAGIC: An Agent based Multimedia Retrieval System
IEEE Transactions on Knowledge and Data Engineering, **16**(9): 1082–1095

Ghosh H, Chaudhury S, Kashyap K, Maiti B. 2007
Ontology Specification and Integration for Multimedia Applications
In *Ontologies in the Context of Information Systems*
Springer

Hammiche S, Benbernou S, Hacid M S, Vakali A. 2004
Semantic Retrieval of Multimedia Data
In *Proceedings of the 2nd ACM international workshop on Multimedia databases*, 36–44

Harit G, Chaudhury S, and Paranjpe J. 2005
Ontology guided access to document images. pp 292–296
In *Proceedings Eighth International Conference on Document Analysis and Recognition, ICDAR-2005*

Harit G, and Chaudhury S. 2006[a]
Clustering in video data: Dealing with heterogeneous semantics of features
Pattern Recognition, **39**(5): 789–811

Harit G and Chaudhury S. 2006[b]
Video Scene Interpretation using perceptual prominence and mise-en-scéne features, pp 571–580 (LNCS 3852). *Asian Conference on Computer Vision (ACCV)* Proceedings published in *Lecture Notes in Computer Science 3852*

Hunter J. 2003
Enhancing the semantic interoperability of multimedia through a core ontology
IEEE Transactions on Circuits and Systems for Video Technology, **13**(1): 49–58

Internet Archive. 2008
Internet Archive
Details available at <http://www.archive.org/index.php> last accessed on 5 September 2008

IGNCA (Indira Gandhi National Center for Arts). 2008
Kalasampada—Digital Library: Resources of Indian Cultural Heritage
Details available at <http://www.ignca.nic.in/dlrich.html> last accessed on 5 September2008

Naphade M R and Huang T S. 2001
A Probabilistic Framework for Semantic Video Indexing, Filtering and Retrieval
IEEE Multimedia Magazine, **8**(1): 141–151

Petridis K, Bloehdorn S, Saathoff C, Simou N, Tzouvaras V, Handschuh S, Kompatsiaris Y, Avrithis Y, Staab S, Strintzis M G. 2005
Knowledge representation and semantic annotation for multimedia analysis and reasoning
IEE Proceedings on Vision, Image and Signal Processing, **153**(3): 255–262

Scottish Distributed Digital Library. 2008
Details available at <http://scone.strath.ac.uk/sddl/index.cfm> last accessed on 5 September 2008

Tsinaraki C, Polydoros P, Kazasis F, Christodoulakis S. 2005
Ontology-based Semantic Indexing for MPEG-7 and TV-Anytime Audiovisual Content
Special issue of Multimedia Tools and Application, Journal on Video Segmentation for Semantic Annotation and Transcoding, **26**, 299–325

W3C. 2001
Web Services Description Language (WSDL) 1.1
Details available at <http://www.w3.org/TR/wsdl/> last accessed on 5 September 2008

W3C. 2004
Web Services Architecture: W3C Working Group Note
Details available at <http://www.w3.org/TR/ws-arch/> last accessed on 5 September 2008

Wattamwar S and Ghosh H. 2008a
Spatio-Temporal Query for Multimedia Database
Workshop on Multimedia Semantics. ACM Multimedia Conference 2008

Wattamwar S, Mishra S and Ghosh H. 2008b
Multimedia Explorer: Content based Multimedia Exploration
IEEE TenCon 2008

Wenzel C and Maus H. 2001
Leveraging corporate context within knowledge-based document analysis and understanding *International Journal of Document Analysis and Recognition (IJDAR)*, **3**:248–260