Multimedia in Cultural Heritage Collections: A Model and Applications*

Cristina Ribeiro[1,2], Gabriel David[1,2], and Catalin Calistru[1,2]

[1] Faculdade de Engenharia, Universidade do Porto
[2] INESC—Porto
Rua Dr. Roberto Frias s/n, 4200-465 Porto, Portugal
{mcr,gtd,catalin}@fe.up.pt

Abstract. The paper presents a multimedia database model accounting for the representation of documents, collections and the associated metadata. Appropriate structures are provided for descriptive metadata and for metadata resulting from automatic content analysis. The model is based on the identification and unification of the main concepts in the archival standards and the audiovisual area.

The main features of the model, designed to support multimedia database applications, are the integration of descriptive and content analysis metadata, the association of metadata to collections as well as to items, the extensibility with respect to the inclusion of new descriptors and the support to several retrieval modes. The MetaMedia application development platform, based on the model, has been used to support the construction of a historic documentation collection where a common web interface provides collection administrators, metadata creators and visitors a multi-faceted view of the repository.

Keywords: Models for multimedia repositories, information retrieval, cultural heritage applications.

1 Introduction

The organization of large digital repositories has required specialists from diverse areas to explore aspects that range from the conceptual aspects of digital collections as information systems [1] to the development of operational platforms to support the organization and access to digital collections [2,3,4]. The success of text-based retrieval has raised the expectations of users concerning the possibilities of search on multimedia collections. Many existing cultural heritage repositories are better viewed as multimedia, and current digital objects are more and more so.

To deal with multimedia collections, it is necessary to handle the content itself, which may require specific storage and presentation devices, and to manage the associated metadata that may be of different nature and be generated according

* Supported by FCT under project POSC/EIA/61109/2004 (DOMIR).

D.H.-L. Goh et al. (Eds.): ICADL 2007, LNCS 4822, pp. 186–195, 2007.

to a variety of standards. In cultural heritage collections, where objects are typically subject to some kind of expert analysis, an operational system must satisfy two goals. The first is to allow rigorous descriptive metadata to be handled and associated to items. The second is to offer content-based search using state-of-the-art technologies.

The paper presents a multimedia database model accounting for the representation of documents, collections and the associated metadata. The model is based on the identification of the main concepts in the archival standards as well as those from the audiovisual area. The main features of the model are the integration of descriptive and content analysis metadata, the association of metadata to collections as well as to items, the extensibility with respect to the inclusion of new descriptors and the support to several retrieval modes.

The model has been designed to support multimedia database applications, and its scope is distinct from the models underlying current standards and from the reference models adopted in several communities. The former [5,6,7], due to their specificity, prescribe strict definitions and formats for the elements of a description. The latter [8,9,10] are more complex models encompassing the objects and their processes in an organization. A model designed to support a multimedia database can be restricted to a set of core features and still allow the incorporation of data from different standards and support the storage and retrieval of individual objects and collections.

The paper is organized as follows. Section 2 introduces the concepts adopted to describe and manage multimedia items. Section 3 outlines the proposed model. The following sections present the MetaMedia application development platform, the case study and the retrieval methods, along with a preliminary evaluation of the proposed framework.

2 Metadata and Standards

The generalized production of digital documents brings new requirements to the activity of librarians and archivists coming from the nature of the documents—increasingly structured and in different media, and the target audiences—ranging from scientists and professionals to unspecialized users.

Metadata covers aspects of media such as its description, content analysis, technical details, terms of use and administrative aspects and it can be automatically generated or manually associated to the documents. There is currently a great interest in automating the metadata extraction process. Automatic analysis of text such as performed by indexing tools can be regarded as an automatic metadata generation process. With visual content, it is also possible to automatically generate descriptors for features such as color or texture.

2.1 Standards

Standards tend to address different aspects of the digital content, the ones most relevant for the intended use of the items in the respective community. While

it may be essential for a library to have detailed information on a scientific journal, title and author descriptors may be enough for documents in a web site; a film distribution corporation may require media- and genre-specific descriptors to provide search on the movie catalog, but the same items when used by a broadcaster require information on the days and times when they will be programmed.

In the library area, there are well-established standards that support applications and metadata sharing [5,11]. Recent work concerns conformance to the XML language syntax, inclusion of sound and image documents, stronger networking.

ISAD and ISAAR [6] are standards for archival description. Their basic principles are multilevel structure and uniform description, and they handle both the records and the people and organizations involved in their creation.

Dublin Core (DC) [12] has appeared to solve the problem of the lack of description of documents on the web. It consists of a set of basic descriptors such as title, creator and date, intentionally kept at a basic non-specialist level. DC is being widely adopted as part of other web-related standards, such as those for the Semantic Web initiatives [13].

MPEG-7 [7] comes from the audiovisual signal processing community and aims at creating metadata for complex multimedia items. The emphasis is on descriptors that can be automatically extracted from audiovisual content, leaving descriptive metadata for other standards such as DC. MPEG-21 [14] originated in the same community and concerns metadata for handling the multimedia delivery chain rather than item or collection descriptions.

3 Multimedia Database Model

The goal of adopting a multimedia database model is twofold. A model can abstract the underlying technologies and therefore allow an application to be maintained as technology evolves. On the other hand, metadata production is expensive and repository managers want to import it whenever available, requiring a model for representing the items.

One thing to consider is whether it is actually necessary to design a model, as most initiatives on standards already rely on their own models and metadata reference models are also available in several domains [8,9]. The fact is a model can only be useful if it treats data at the convenient abstraction level. Following the model for one standard would lead to a specialization in its application domain, whereas reference models have a much broader scope than required for the basic representation and retrieval tasks. We have therefore chosen to design a compact model capturing concepts from several relevant standards and amenable for implementation. The model is centered on the integration of descriptive and content analysis metadata of multimedia objects, to support automatic indexing and retrieval tasks.

In the sequel we will first present the core concepts of the MetaMedia model and then an outline of the data model with the relevant relationships.

3.1 The Concepts

The model is organized around four main principles. The first one is that multimedia objects are usually organized in a *part-of hierarchy*. To each level one can associate a set of attributes characterizing the corresponding set of items. These attributes typically cover aspects related to the creation context of the multimedia object, its format, support and access conditions.

The second principle is that of *uniform description*, whereby the same set of attributes is used for an individual object, for composite objects and for sets of related objects. This principle has been followed in the standards for archival description such as ISAD [6] and is very useful in the representation of large collections: metadata is frequently available for sets of items rather than items, and inheritance can make it useful further down the hierarchy.

The third principle applies to actual multimedia data. It is concerned with the internal structure of individual multimedia objects. The actual multimedia object is stored in one or more *segments* which are parts of multimedia items.

The contents of each segment can be analysed by appropriate tools, which generate specialized *descriptors*. For example, the video track segment of a specific multimedia item may have associated motion activity and color descriptors, while the audio track segment of the same item may be connected to a melody contour descriptor. The fourth principle states that the descriptor resulting from the analysis of some feature of a segment is expressed as an XML file of an appropriate format.

According to the principles just outlined, four main concepts were selected. The first one, the Description Unit (DU), is already present in archival description [6], corresponds to the concept of Digital Item in the audiovisual standards and captures the notion of an object or collection of objects with an associated context.

DUs are organized in hierarchies that may have various topologies and different semantics for their levels. Such hierarchies can be created for new collections and can be extracted from existing ones. In both cases they capture the nature of the datasets. A Scheme that defines the possible levels, their semantics and their interconnections is required. Figure 1 shows a sample from a part-of hierarchy and the corresponding scheme.

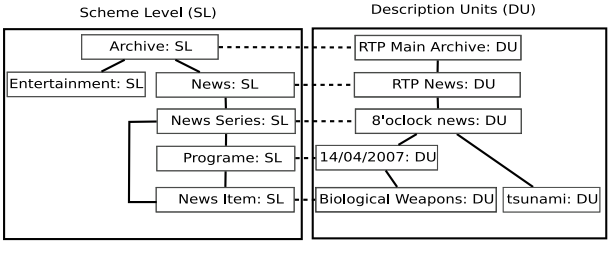

Fig. 1. Scheme Instance Example

The third concept is the Segment, following the MPEG-7 vocabulary, and captures the notion of some part of an actual multimedia item, such as a video sequence that is reused in a new documentary work. A segment has no context of its own, getting it from the DU of the object it belongs to.

The fourth concept is that of a Descriptor. The sense in which Descriptor is used is the one established by the MPEG-7 standard—a representation of a feature [7]. In the model, a Descriptor is regarded as the result of analysing a Segment. An image Segment, for example, can be associated to its corresponding instances of the *DominantColor* and *NumberOfFaces* descriptors, a video Segment can be associated to its *MotionActivity* descriptor and an audio Segment to its *MelodyContour*.

3.2 The Data Model

Figure 2 shows a simplified version of the data model. The concepts in the model are associated with the main classes. Control of the hierarchy is provided by the Scheme Level class. Each Description Unit is of a specified level; the structure of the levels, omitted in the simplified model, allows the application development platform to enforce creation and structure of the DUs according to the repository schema. In Figure 1, for instance, a Description Unit at the level of News Item can be a direct descendant of instances of Program or News Series, but not of News or Archive. Attributes such as title, author, date and copyright apply at the fine-grained level of the document as well as at the coarser grain of the whole collection. They are therefore captured as attributes in the Description Unit and appear uniformly at all levels.

The Segment class embodies the corresponding concept, modeling object fragments that belong to items in the Description Units. The association between Description Units and Segments is visible as the Contents class. Segments are further specialized as text, image, video and audio.

Descriptors such as *MotionActivity*, *DominantColor* and *MelodyContour* are likely to be used only for specific kinds of Segments. The Descriptor class represents them and the association of segments to descriptors is modeled with the Descriptor Instance class.

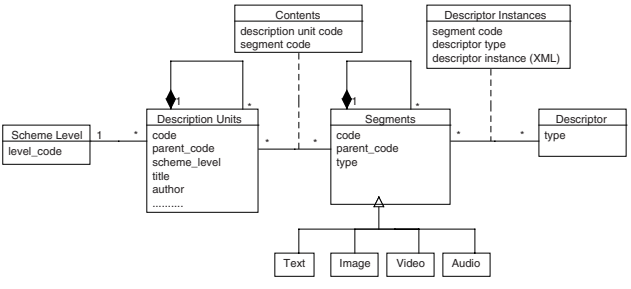

Fig. 2. The MetaMedia Model

In the model, the DU and Segment classes are very similar in structure: both offer a hierarchical organization of their objects. The main distinction lies in the nature of the objects and in the kind of associated metadata. An instance of Description Unit captures a document for which a well-established description is available, and which has been related to other documents according to the repository hierarchy. An instance of segment is appropriate for representing a video frame for an object whose description is in the associated DU and for which some automatic low-level descriptors have been produced. The part-of structure for the segments has no predefined structure and is intended to follow the granularity of the existing analysis tools.

4 MetaMedia — A Multimedia Database Platform

The current MetaMedia model has evolved from an early version of the model [15,16]. Representation of descriptors in XML allows the incremental addition of descriptors with arbitrary structure. This feature is important both for representing text annotations according to specified vocabularies and for dealing with multimedia items. Processing visual and audio information is a very active area of research and new descriptors become available frequently [17].

Having a model with a compact set of concepts has allowed the development of a software platform where object collections, object relations, high-level metadata and diverse low-level descriptors can be managed. The platform has been designed as a web portal supported on a database system, and is currently under test using several kinds of collections.

A collection interface built on the MetaMedia platform gives access to the structure of Description Units, but also to diverse associated segments. A collection of text documents may have extracted text—a segment for each document, and automatically extracted descriptors for the text—content annotations based on a controlled vocabulary. A video collection may have multidimensional descriptors such as *Color* or *Motion Activity*, automatically obtained for some of the video segments. Each kind of descriptor will be used for content-based access to the objects in the collection.

Content-based retrieval of textual documents has succeeded in providing useful answers to common queries based on automatic indexing. Content-based retrieval of audiovisual materials has to deal with the "semantic gap" [18] separating the low semantic level of automatically extracted descriptors from the high-level concepts people use in search. Many experimental systems explore content-based retrieval of multimedia objects [17,18]. The approach in MetaMedia is to allow search using the high-level concepts to be combined with search based on object similarity. There are no systems with these combined features in current production.

5 The "Terra de Santa Maria" Documentation Center

The "Terra de Santa Maria" historic documentation center [19] has been a case study for the MetaMedia platform. The collection is a virtual archive of medieval

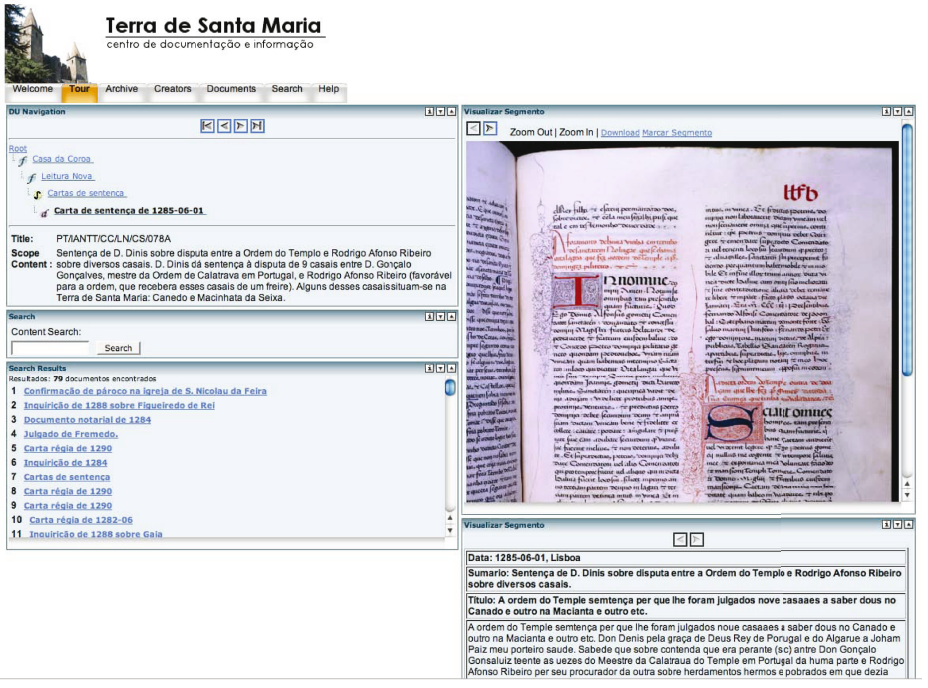

Fig. 3. Interface of the Historic Archive

documents, for which there are transcriptions in either Latin or archaic Portuguese. Information for a document in the repository includes three parts: the digitized image, the document transcription and the archival description according to the ISAD/ISAAR standards.

The interface is available in English and in Portuguese, and the archival descriptions have been generated in Portuguese. There are several views on the document, intended for different kinds of users.

On the *Archive* tab, it is possible to browse the structure of the archive, designed by archivists, and to edit the descriptions. The *Creators* tab has detailed information on the creators for the current unit. Under the *Document* tab users may explore the contents of the documents. A medievalist studying one of the parchments uses this mode to observe the digitized image and its transcription side by side and to possibly upload his own analysis of the text.

Specialists need to search on specific items, while casual users require a simple keyword-based retrieval mode. Both are offered, and it is also possible to use the concepts marked up on the document transcriptions for search purposes. Textual segments corresponding to the digital content have been marked using an XML tool. Markup identifies key concepts that would be hard to spot on the original latin documents and is mainly intended for improving retrieval.

6 Collection Building and Management

Interesting documents are essential when we want to offer access to a digital repository, and the available metadata may allow rich views on their contents. A tradeoff exists between the simplicity of the view which is offered on documents and their relations and the detailed access to the collection building tools. Having a compact set of concepts underlying the model helps to put the collection manager in control of the relevant aspects of the collection and to automate the building process.

In the MetaMedia platform there is one task strictly reserved for the collection administrator: the design of the repository hierarchic structure. The part-of relationship exists in traditional archives due to their provenance-based structure, but can also be found in most large collections, where it is useful to account for the organization of collections and sub-collections and deal with the increasing complexity of digital items.

Based on a chosen hierarchy, items and their metadata are added. Descriptive metadata is available both at document and at collection level, and it is possible to add it at any point in the collection production. Segments in the form of text, image or video may also be associated to items at any point. Taking advantage of this information in search requires a rebuild of the low-level indexes and automatic descriptors. A collection is ready for presentation when the automatic addition of extracted descriptors and the indexing of the textual and non-textual content have been performed.

7 Search and Retrieval

A multimedia repository must offer several retrieval modes. The user may search on textual content for textual segments of the documents, on the structured contextual metadata available as descriptors in the Description Units, and on the visual features for image/video segments of the items.

Query by keyword is the most straightforward and requires a full-text indexing system. The platform uses the Apache Lucene search technology [20], configured to index selected parts of the available textual information: the textual content itself and parts of the descriptive information. A structured query interface is intended for specialized users, who are aware of the meaning of the descriptive metadata in the Description Units. Search on the contextual metadata is handled by the built-in indexes of the relational database management system.

Query by visual features is integrated with the keyword search. In a collection where documents have textual segments, keyword search provides an initial answer which is expanded by visual and audio similarity. A collection with just image or video objects can also be searched with textual queries, provided that some minimal concept annotation is present. Concepts in the query are extracted and matched against the annotations, resulting in an initial set of items. Image and audio similarity are then used to expand the answer and a relevance feedback interface is offered to refine the query. The “BitMatrix” multidimensional

indexing technique [21] is used to ease the computation of image similarity and to allow the tuning of the set of descriptors to the nature of the collection.

When collections have substantial structure, browsing can be an effective way of guiding a retrieval process. The MetaMedia platform keeps the context of items being visited, and it is therefore possible to browse the hierarchy, locate an interesting sub-collection, view its description, travel down to a document and analyze its visual or textual content. From a selected image or video segment it is also possible to start a new search using low-level similarity.

8 Conclusions

The representations used by libraries and archives are being extended to deal with audiovisual content. Object descriptions must capture their features and also support the navigation and search by unspecialized users. Many standards have emerged, addressing common requirements and specificities of some content types and application domains.

In this work we have identified the main concepts underlying the standards and built a compact model suitable for implementation in a relational database system with associated repositories for multimedia items. The model allows structured descriptors to deal with the fine-grained features of digital items on various domains. A web platform is used to build repositories, to associate metadata to the items and to search the collections.

The contributions or the paper are: the analysis of existing standards from the point of view of an operational multimedia database system; a multimedia database model; a platform for managing, building and accessing digital repositories; and a search interface integrating descriptive and content metadata where high- and low-level descriptors are interleaved.

A case study shows the appropriateness of the model in a historic documentation center, an application domain where diverse description and content analysis metadata is available. The model has proved robust and the configurable hierarchy has been useful both for representing the structure of the repository and for making it accessible to the public.

The MetaMedia platform is currently being used for searching video collections with no descriptive metadata. The video retrieval process is being prepared for a retrieval evaluation process. Several open issues remain, namely in the search with automatically extracted descriptors, the collaborative environment for uploading segment descriptors and the user interfaces for complex items.

References

1. Gonçalves, M.A., Fox, E.A., Watson, L.T., Kipp, N.A.: Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries. ACM Trans. Inf. Syst. 22, 270–312 (2004)
2. DSpace: The DSpace digital repository system (2007), `http://dspace.org/`
3. Lagoze, C., Payette, S., Shin, E., Wilper, C.: Fedora: an architecture for complex objects and their relationships. Int. J. on Digital Libraries 6, 124–138 (2006)

4. Witten, I.H., Bainbridge, D.: Building digital library collections with Greenstone.. In: JCDL, p. 425 (2005)
5. Library of Congress: MAchine-Readable Cataloguing (MARC) (2007), http://www.loc.gov/marc/
6. ISAD(G): General International Standard Archival Description, Second edition (1999), http://www.ica.org/
7. Sanchez, J.M.M., Koenen, R., Pereira, F.: MPEG-7: The Generic Multimedia Content Description Standard, Part 1. IEEE MultiMedia 9, 78–87 (2002)
8. International Council of Museums: CIDOC Conceptual Reference Model (2006), http://cidoc.ics.forth.gr/
9. DELOS Network of Excellence on Digital Libraries: A Reference Model for Digital Library Management Systems (2007), http://www.delos.info/
10. Hunter, J.: Combining the CIDOC CRM and MPEG-7 to Describe Multimedia in Museums. In: Proceedings of MW 2002: Museums and the Web, Archives and Museum Informatics (2002)
11. Library of Congress: Metadata Encoding and Transmission Standard (METS) (2007), http://www.loc.gov/standards/mets/
12. Dublin Core Requirements Group: Dublin Core Metadata Initiative (2007), http://dublincore.org/
13. W3C Consortium: Semantic Web (2007), http://www.w3.org/2001/sw/
14. Burnett, I., Van de Walle, R., Hill, K., Bormans, J., Pereira, F.: MPEG-21: Goals and Achievements. IEEE MultiMedia 10, 60–70 (2003)
15. Ribeiro, C., David, G.: A Metadata Model for Multimedia Databases. In: Proceedings of ICHIM 2001: International Cultural Heritage Informatics Meeting, Archives and Museum Informatics (2001)
16. Ribeiro, C., David, G., Calistru, C.: A Multimedia Database Workbench for Content and Context Retrieval. In: MMSP IEEE Workshop, IEEE Computer Society Press, Los Alamitos (2004)
17. Lew, M.S., Sebe, N., Djeraba, C., Jain, R.: Content-based multimedia information retrieval: State of the art and challenges. ACM Trans. Multimedia Comput. Commun. Appl. 2, 1–19 (2006)
18. Zhao, R., Grosky, W.I.: Narrowing the Semantic Gap—Improved Text-Based Web Document Retrieval Using Visual Features. IEEE Transactions on Multimedia 4, 189–200 (2002)
19. Comissão de Vigilância do Castelo de Santa Maria da Feira.: Centro de Documentação da Terra de Santa Maria (2007), http://www.castelodafeira.pt/
20. Apache Software Foundation: Apache Lucene 2.2 (2007), http://lucene.apache.org/
21. Calistru, C., Ribeiro, C., David, G.: Multidimensional Descriptor Indexing: Exploring the BitMatrix.. In: Sundaram, H., Naphade, M., Smith, J.R., Rui, Y. (eds.) CIVR 2006. LNCS, vol. 4071, pp. 401–410. Springer, Heidelberg (2006)