Searching, browsing, and editing metadata for Audio/Video documents

Giuseppe Amato
CLAUDIO GENNARO
Pasquale Savino

ISTI – CNR
Pisa
Italy
claudio.gennaro@isti.cnr.it

Outline

◆ Manual indexing of A/V documents
  ▪ Introduction
  ▪ IBM MPEG7 editor tool (demo)
  ▪ Tecmath example
  ▪ The echo model (restricted version)
    • The echo Metadata editor

◆ Browsing the echo Metadata
  ▪ Demonstration of an A/V Digital Library

◆ Echo Editor Architecture (?)
◆ Echo Editor demo
Editing Audio/Video metadata

Metadata classification

- Bibliographic
  - Title
  - Author
  - Collocation/Storage information

- Temporal/Spatial decomposition
  - Audio: noise, speech, etc.
  - Video: scenes, objects, etc.
  - Map: cities, rivers, etc.
Video decomposition

◆ Mono-Dimensional: scenes
◆ Multi-Dimensional: stratum of scenes, of audio, of transcript, etc.
◆ Hierarchical: scenes, shots, etc.
◆ View based
Segments representation

- Simple mono-dimensional

Multi-dimensional (strata)
Segments representation(3)

- Hierarchical

Importance of the analysis tools

- More sophisticated software can aid the cataloguers during the editing phase:
  - Intelligent tool of cut detection, which do not produce masses of scenes (tunable: camera movement, fade, etc.)
  - Face and object detection tools are very useful
  - Audio analysis and speech recognition
Model or meta-model?

- Sometimes editing tools allow the user to modify the model
- IBM editor allows to modify the lexicon, or
- the tecmath editor allows to add/remove strata
- There must be a separate tool for maintaining the metadata model
  - Confusion
  - Inconsistence, adaptation of other interface (e.g., the retrieval interface)
Usability

- This issue is very important in audio/video metadata editing tools
- Silver editor of the Informedia project of the Carnegie Mellon University faces this problem
- This tool is more a video editing tool rather than a metadata editor, however the problem faced are similar

VideoAnnEx Annotation Tool
The VideoAnnEx annotation tool assists authors in the task of annotating video sequences with MPEG-7 metadata.

Each shot in the video sequence can be annotated with static scene descriptions, key object descriptions, event descriptions, and other lexicon sets.

The annotated descriptions are associated with each video shot and are stored as MPEG-7 descriptions in an output XML file.

The annotation tool also allows customized lexicons to be created, saved, downloaded, and updated.
Download site


Editing metadata with the standard strata-based metadata editor of tecmath

* It is the standard metadata editor provided with the retrieval interface of the tecmath commercial system.
Keyframes

Videosegments (strata)
Scenes detected

Audio segments
Il numero delle auto a sosta è più che triplicato nel dopoguerra, anche se questa è stata una progressione parallela all'incremento della popolazione. Il problema è stato molto risolto con l'introduzione diremove e le disposizioni molteplici della circolazione stessa. Questa è unica e....
New stratum name
Editing text associated with detected scenes

Write the word “bus” in the scene...
The Echo project example

Motivations

- This model originated from our experience in the ECHO project (European CHronicle On-line)
  - ECHO is an EC funded IST project
  - ECHO aims at providing
    - remote access to collection of historical documentary audio-video resources
    - a software infrastructure to support digital video archives
    - an extensible and interoperable architecture
The system architecture

Application

Logic

Other services
(e.g. automatic
processing)

Metadata editor

Speech recognition

Metadata DB

Audio/Video DB

User Interface

Logic

Typical three-tier architecture

Other services
(e.g. automatic
processing)

XML

Application Logic

Audio/Video DB

Metadata DB

JDBC/ODBC

DLIB Project - November 27-28, 2003 - Pisa, Italy
The approach

The metadata model uses and extends the IFLA-FRBR methodology:
- Entities that describe different aspects of a resource
  - WORK
  - EXPRESSION
  - MANIFESTATION
  - ITEM

The metadata model

- Bibliographic metadata
- Time/space related (multimedia) metadata

It provides the general intellectual or artistic view of the document.

They concern to the nature catalogued object. In particular in our model we focused on the time/spatial aspect of the multimedia object, i.e., how the audio/video object are divided in scene, shots, etc. (White Box).

E.g., if the data were geographic maps, and the metadata would be points (cities), regions (lakes, rivers).
The ECHO Model

**WORK**

**AVDOCUMENT**

Attributes:
- Title
- EventDate
- Director
- Censorship
- Description

---

**EXPRESSION**

**TRANSCRIPT**

Attributes:
- AudioLanguage
- Frequency
- Type

**PartOf**

FollowedBy

Attributes:
- Transcript
- SpeakerID
- Gender
- SpeakerLanguage

---

Attributes:
- VersionTitle
- Duration

---

Attributes:
- Version
- Duration

---

Attributes:
- SubtitleLanguage
- ImageDescription
- VideoAbstract
- KeyFrame
- Faces
- Objects
- CameraMovement

---

Attributes:
- VideoLanguage
- Frequency
- Type

DLIB Project - November 27-28, 2003 - Pisa, Italy
The ECHO Model (cont.)

MANIFESTATION

MEDIA

Attributes:
- Format
- Size
- ...

Available As

ManifestedBy

Available As

SyncronisedWith

The ECHO Model (cont.)

ITEM

STORAGE

Attributes:
- Collocation
- Provider
- StorageID
- PublicAccess

Available As
Regia: A Metadata Editor for A/V Digital Libraries

How the AVDocument is stored

- Each instance of the entities (objects) are stored in an individual xml file in the disk or in the database
**Terminology: The GUID**

- **GUID**: is a unique identifier of an object in the model. However, the type of GUID changes.
- When the object documents are stored in the disk (FileSystem) the GUID is the complete path filename of the XML document representing the object.
- When the object is stored in the Database it is the URN of the document.

**Terminology: The AVDocument**

Throughout this presentation we will use the term AVDocument for referring to a whole document of the editor which is an instance of the whole entity object model above presented. In order to avoid misunderstanding with term AVDocument indicating the entity of the model and the editor document when we refer to the AVDocument we call it “AVDocument root”.
Opening a AVDocument from the filesystem

- menu File → Open and browse and select a root xml file of the AVDocument

Opening a AVDocument from the filesystem (2)

- It is possible to open an entity of the AVDocument selecting a file xml in the subdirectory of the AVDocument. In this case Regia will open a the whole document and will open also the edit dialog of the selected object.
Opening a document from the Database

- On the ECHO retrieval interface click on the title of one of the AVDocuments retrieved
Browsing/Editing the Bibliographic section AVDocument

Each entity is associated to a different icon:

- Root of the AVDocument
- Version
- Media
- Storage
The bibliographic metadata structure

Editing the metadata
Some type of metadata fields do not have null values:

- Boolean (true/false)
- Date
- Closed Lists
Delete/Add objects

◆ **Deleting an object and its children**
  - By selecting an object icon and typing the “Del” button on the keyboard it is possible to delete an object and its children.

◆ **Adding a new child entity**
  - It is possible to add a new child object by selecting its parent, clicking on the right button of the mouse, and clicking on “Add child...”. It will appear an input box asking for the title of the new object. Note that, for the Storage entity the title corresponds to the “collocation” field of its metadata.

Multimedia section of AVDocument
Expression Tool

Browse the Expression
Create, update and delete

◆ Create
  ▪ move the cursor of the slider and place it at the start time of the shot you want to create, push the button “|”. The timecode and a small line on the timeline start blinking. Move the cursor and place it at the end time of the new segment, push again the button “|”.
  ▪ Push the button “|--|”. A dialog box will allow you to select the start/end timecodes boundaries of the new segment.

◆ Modify
  ▪ Push the button “|>” (“<|”). This will change the end (begin) of the selected segment to the begin (end) to the (preceding) succeeding segment.
  ▪ Push the button “<->”. A dialog box will allow modifying the start/end timecodes boundaries of the selected segment.
  ▪ By selecting the segments by means of the mouse button and pressing shift button, it is possible to select more than one segment. The selected segments are highlighted by means of a thicker line on the sliding bar, and their keyframes are contained in a coloured frame. By pushing the button “-+-” all the segmented will be merged together. Note that, the new merged segmented will hierarch the textual metadata of the selected segment that appeared in green.

◆ Delete
  ▪ Select a segment and push the button “X” or the key “Del” of the keyboard.
Updating the keyframes

- Clicking on the button “K” the keyframe of the current video is updated by the frame currently shown.

Creating a new AVDocument

- In order to create a new AVDocument from scratch, click on the menu item File→New. A new AVDocument will be created. Proceed creating the other entities of the AVDocument.
Cross Language features

◆ Some of the metadata fields of the root of the AVDocument are treated as Cross Language by the editor. The metadata fields involved are:
  - Themes
  - Subthemes
  - Thematic Keywords
Browsing the metadata: the Echo Retrieval Interface

Logging
The retrieval Interface presentation
Retrieval tools interfaces

Metadata of the Selected document

List of documents retrieved

Retrieval Interfaces: search for documents using different techniques
Free text search interface
Free text to search
Restrict the search on a specific entity
period of creation time
Restrict the search on a specific archive

Strategies of sorting results
Order of sorting

Number of results
Fielded search interface
Logical connector among fields

Field choice
Multilingual classification
The documents in the collection are classified on the basis of Genres, and Themes, Subthemes, and Thematic keywords. The label of these categories are translated in 5 language (English, Italian, Dutch, French, and German).

A document can be associated to one theme, one subtheme, and one or more thematic keywords.
Crosslanguage retrieval interface

Collection

Theme selection

Free text to search (any language)
Image similarity search
Browsing the document structure
metadata

Result set

Word “train” found
The multimedia metadata
Video player

Scenes detected

Keyframes

Duration

Video segmentation

Video player
The fielded search
The transcript of a video can be in Italian, French, German, or Dutch.

The CL retrieval allows to search the transcripts using any of these 4 languages, and the English language.
Search for the Italian word “ragazza” (girl) found in transcripts.

Search for the Italian word “ragazzo” (boy).

Search for the English word “boy”.

Italian word “ragazzi” (boys) found in transcripts.
The MultiLingual Classification
The Image Similarity search
The End

Regia architecture

Claudio Gennaro
gennaro@iei.pi.cnr.it
**The editor behavior**

Search interface

- GIUDs (XML)
- mpeg file GUID (all)

**The Editor architecture**

- Interface classes
- Data classes
- Work
  - AVDoc
  - Video
  - Audio
- Expression
  - Transcript
- Manif.
  - Media
- Item
  - Storage
- DOM
  - xml strings
- xml files
- schema
- Corba
- Data Manager
How the xml schema is used

- It defines the schema of the echo xml “documents” (AVDocument, Video, etc.);
- It is used as configuration file for the editor:
  - When the editor opens an xml document it reads the schema;
  - In this way it is possible to modify the schema of the document;
  - It is also possible to say to the editor that a field is readonly or to be ignored.

Editing the xml schema

- The editor recognizes the following data-types:
  - xsd:string, xsd:integer, xsd:date, xsd:float, xsd:boolean (predefined in XSD)
  - SetOfString (a simple list of string items)
  - Close Lists
  - Multimedia type (objects, faces, etc.)
Example of xml schema

```xml
<?xml version="1.0" encoding="UTF-8"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2000/10/XMLSchema"
    elementFormDefault="qualified">

    <xsd:element name="AVDocument">
        <xsd:complexType>
            <xsd:sequence>
                <xsd:element name="Title" type="xsd:string">
                    <xsd:annotation>
                        <xsd:documentation>Title of the av document (original title if known otherwise assigned)</xsd:documentation>
                        <xsd:appinfo>readonly</xsd:appinfo>
                    </xsd:annotation>
                </xsd:element>
                ... 
                <xsd:element/>
            </xsd:sequence>
        </xsd:complexType>
    </xsd:element>

</xsd:schema>
```

This defines the AVDocument, which is a sequence of "fields". This defines the first field of AVDocument and so on...

Whose name is "Title" and whose type is string. This is an annotation for the human reader, which is also showed as tip on the editor interface.

I can define a close list as following:

```xml
<xsd:simpleType name="CutType">
    <xsd:restriction base="xsd:string">
        <xsd:enumeration value="HardCut"/>
        <xsd:enumeration value="FadeIn"/>
        <xsd:enumeration value="FadeOut"/>
        <xsd:enumeration value="FadeInFadeOut"/>
    </xsd:restriction>
</xsd:simpleType>
```

The list of items: HardCut, FadeIn, FadeOut, FadeInFadeOut

Which appears as combo box in the editor interface
Example of xml document

```
<AVDocument xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="C:\Documenti\Schema\echo_schema.xsd">
  <Title>La vita di Fellini</Title>
  <SeriesTitle>I don't know</SeriesTitle>
  <SeriesNumber>323224</SeriesNumber>
  <Genre>Documentario</Genre>
  <Description>Racconta la vita del grande regista Fellini</Description>
  <PersonNames>
    <StringItem>Marcello Mastroianni</StringItem>
    <StringItem>Giulietta Masini</StringItem>
  </PersonNames>
  <Themes>Biografie</Themes>
  <Subthemes>Registi</Subthemes>
  <ThematicKeywords/>
  <EnglishAbstract>This is a documentary film describing the life of the famous Italian director Fellini</EnglishAbstract>
  <EventDate>2001-05-11</EventDate>
  <DescriptionLanguage>Italian</DescriptionLanguage>
  <ProductionDate>1999-06-12</ProductionDate>
  <ExpressedBy>
    <StringItem>FE4535A32345</StringItem>
    <StringItem>FE4535A32346</StringItem>
  </ExpressedBy>
</AVDocument>
```

Editor Demo