Int J Digit Libr (2004) 4: 3–7 / Digital Object Identifier (DOI) 10.1007/s00799-003-0062-8

ECHO: a digital library for historical film archives

Pasquale Savino, Carol Peters

ISTI-CNR, Area di Ricerca di Pisa, 56124 Pisa, Italy
e-mail: {savino, peters}@isti.cnr.it

Published online: 27 July 2004 – © Springer-Verlag 2004

Abstract. Easy access to large information collections is of great importance in many aspects of everyday life. However, limitations in information and communication technologies have so far prevented the average person from taking much advantage of existing resources. Historical documentaries held by national archives constitute some of the most precious yet least accessible cultural information. The ECHO project has facilitated accessibility to this type of precious information by developing a digital library (DL) service for historical films belonging to large national audiovisual archives.

Keywords: Digital library – Audio/Video – Metadata – Metadata editor – Information retrieval

1 Introduction

The objective of the ECHO project has been to develop a DL service for historical films owned by large national audiovisual archives.[1] Actually being able to see and hear an account of a historical event, filmed in the original context, is very different from reading about it. The ECHO services allow users to search and access these documentary film collections and exploit the content for their own particular requirements, whether commercial, educational, leisure, or whatever.

The project involved a number of European institutions holding or managing unique collections of documentary films dating from the beginning of the century until the 1970s: Istituto Luce, Italy; Institut Nationale Audiovisuel, France; Netherlands Audiovisual Archive; and Memoriav, Switzerland.[2] These collections are of great value as they document various aspects of life in Europe (social, cultural, political, economic) during this time period. The set of services implemented by ECHO thus provides users with access to significant portions of their cultural heritage, which would otherwise be almost inaccessible. In addition, users can compare the way in which an event or phenomenon is documented in their own country with how it is reported in others, or they can investigate how different countries have documented a particular historical period of their life, etc. This means that ECHO services have to operate across linguistic, cultural, and national boundaries while respecting the requirements of international standards.

This paper briefly presents the ECHO system, looking at its functionality, outlining the system architecture currently implemented, describing the metadata editor, and listing the advanced functionality included in the final prototype.

2 System functionality

The services to be provided by the system were defined on the basis of the results of a user needs analysis performed by the project. A main requirement was to support interoperability over distributed, heterogeneous digital collections and services. Achieving interoperability in the DL setting is facilitated by conformance to an open architecture as well as agreement on items such as formats, data types, and metadata conventions. These issues have already been addressed with varying degrees of success by DLs handling textual collections; the challenge in ECHO was to solve the numerous technical problems that up until now have hindered the inclusion of audiovisual material in a searchable digital environment. The aim has

[1] ECHO was funded by the IST programme of the European Commission under the V Framework; the project began in February 2000 and was completed in March 2003 [2].

[2] Academic partners were ISTI-CNR (coordinator), CNRS-LIMSI, ITC-irst, University of Twente, University of Mannheim, and Carnegie Mellon; industrial partners were Techmath and Eurospider Information Technology.

been to make the film collections available and searchable to as broad a range of users as possible. To achieve this goal, the following components were developed and included in the system.

2.1 Audiovisual metadata model

When the project began, there were no well-defined metadata models for an adequate description of film data. A major effort of the project has been to define a suitable metadata model to represent the audiovisual contents of the archive. The model that has been implemented is an adaptation of the IFLA-FRBR model, a general conceptual framework used to describe heterogeneous digital media resources [4]. The ECHO metadata model thus adapts the well-known IFLA four levels describing different aspects of intellectual endeavour: work, expression, manifestation, and item to include new subentities to describe audiovisual documents. For example, the IFLA work entity has been extended by defining the subentity AVDocument, which contains attributes such as Director, Event, Date, Person, Location, and Description and the expression entity has been extended by defining the version entity that contains specialized attributes like VersionTitle, Duration, etc. The entities of the model are hierarchically ordered from the top level (work) to the bottom (item). A full description of this model can be found in [1].

2.2 Intelligent access

The ECHO system assists the application developer during the indexing and retrieval of audiovisual documentaries. Semiautomatic indexing is supported: the system automatically extracts several items of metadata information such as the scenes composing the video, keyframes that describe each scene, image features describing each keyframe, spoken dialog (automatically transformed into text through a speech recognition process), faces, and specific objects. Later on, the developer can complete the indexing by specifying metadata that cannot be automatically extracted. Search and retrieval via desktop computer and wide area networks is performed by expressing queries on the audio transcript, on the metadata, or by image similarity retrieval. Retrieved films or their abstracts are then presented to the user. By the collaborative interaction of image, speech, and natural language understanding technology, the system compensates for problems of interpretation and search that arise when handling the error-full and ambiguous datasets.

2.3 Multilingual user interface

The ECHO film archives are made up of language-dependent (speech, text) and language-independent (video) media. Thus, although users querying over collections in different languages may not understand the spoken dialog, they can still identify useful documents (or parts of documents) via the images. This has facilitated the implementation of a relatively simple multilingual search interface that can still provide useful functionality. The approach adopted has been to implement online cross-language search tools based on the use of standard metadata formats and mechanisms that provide a mapping between controlled vocabularies agreed between the content providers. Access is provided by local site interfaces in the local languages, but a common user interface in English is also maintained on the project Web site for external access.

2.4 Creating visual summarization

The project has developed techniques to produce visual summaries. The aim is to capture the content and structure of the underlying documentary film in a brief visual abstract. The summary consists of a sequence of moving images, much shorter than the original film, but preserving the essence of the original message. It should provide a good overview of the entire film documentary. The creation of visual summaries is based on the use of a number of video features such as the different scenes recognized, faces, text, objects, and action scenes detected. After this initial analysis, the more relevant clips are determined and assembled to maintain the flow of the story. The abstract is usually set to 8% of the length of the original video, but other values can be specified depending on user and application needs. The video abstracting process is performed offline, after video archiving, since it requires approximately ten times the video duration.

2.5 Security

To make a DL of films possible, the copyright owners must be guaranteed that their property will be protected and that its use will be measured to provide them with appropriate compensation. ECHO thus includes mechanisms that support access control, authentication, security, and privacy.

3 System architecture

The architecture of ECHO consists of three main components: client interface, automatic processor, and middleware (Fig. 1). The client interface is the component directly employed by the user to interact with the system. The automatic processor analyzes the multimedia documents to automatically extract the metadata. The middleware component manages access to data stored in the video and metadata databases for the other two components.

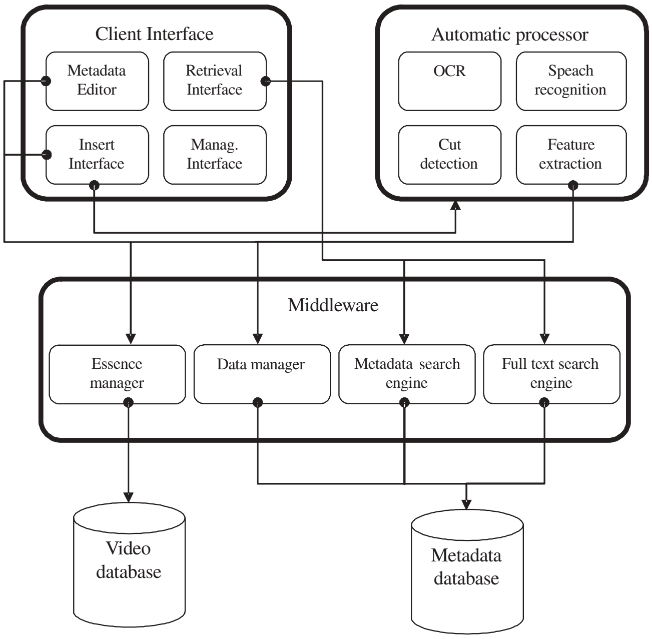


Fig. 1. ECHO system architecture

3.1 Client interface

The client interface is composed of four main modules. The metadata editor allows users to manually edit and review metadata associated with multimedia documents. The user can either edit automatically generated metadata, such as scene boundaries, or add additional metadata manually. The insert interface is used when new documents are inserted. This module interacts with the metadata editor and the automatic processor components that analyze the documents being inserted. The retrieval interface is used to search the system for documents of interest. Various possibilities are offered by this interface. Users can retrieve documents by performing full text retrieval on the transcript or on descriptions associated with documents, or by selecting specific fields of the metadata structure. Finally, the management interface can be used to configure and fine-tune the system.

3.2 Automatic processor

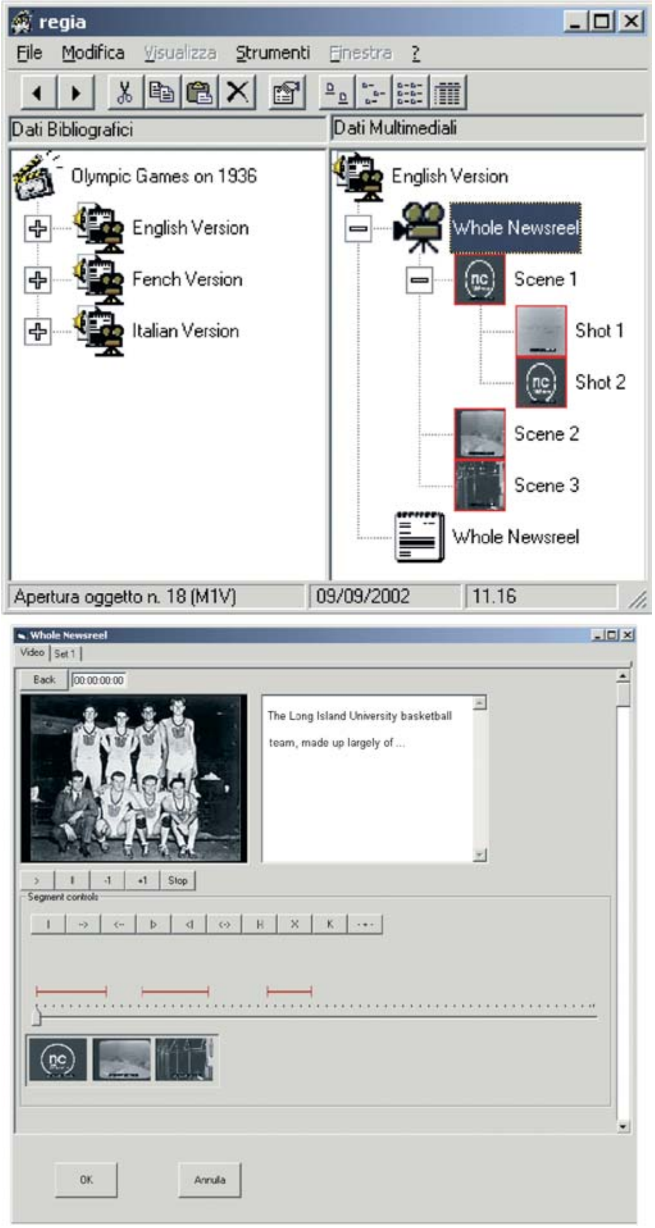
The automatic processor is composed of four main modules, each dedicated to a different automatic processing technique. The speech recognition module is able to generate a transcript corresponding to an audio or audiovisual document. The generated transcript is indexed, and the corresponding document can be retrieved by performing full text retrieval. The cut detection module analyzes a video document and automatically identifies scene changes. In this way, metadata can be associated with specific portions of a document instead of the whole document. The OCR module performs automatic character recognition. Finally, the feature extraction module analyzes multimedia documents to extract physical properties, which can be used to perform similarity retrieval. Typical features extracted are color distribution, texture, shapes, and motion vectors [3]. Higher-level features are faces and specific relevant objects.

3.3 The middleware

This component manages the accesses to the underlying databases: the video database, which physically stores video documents managed by the system, and the metadata database, where all metadata associated with the documents are stored. The middleware component consists of four modules: the essence manager, which handles the access to the video server for the other system modules, e.g., the metadata editor or the automatic processor; the data manager which handles the access to the metadata database; the metadata search engine, which supports document search based on the full metadata content; and the full text search engine, which can be used to search for documents by using textual parts of the metadata. For instance, the descriptions and the transcripts associated with documents are used to perform this type of search. The results of the two search engines can be combined to allow users to perform an integrated complex search.

4 Metadata editor

The metadata editor reflects the complexity of the metadata model. It has been designed to be used by catalogers inserting new audiovisual documents into the archives. The interface of the editor makes it possible to browse an audiovisual document according to a treelike structure. The metadata that belong to different classes included in the model are logically divided into two sets: bibliographic metadata and time/space metadata. This classification is also reflected in the Metadata Editor interface. Figure 2 shows a screenshot of the interface: the window on the top displays a document as a folder-like navigation tool. At the top level of the tree is an icon representing an AVDocument object at the Work level ("Olympic Games in 1936" in our example). In connection with the Work object the editor presents the three main versions that belong to this particular AVDocument. By selecting an icon representing a version (the Italian version in the figure), the media instances of the version can be seen together with the corresponding storage objects. The navigation tool on the left side of the window shows only the main expressions belonging to the documents (i.e., the expression that corresponds to the entire audiovideo document). The editor can browse single versions one at a time by using a second frame on the right side of the window. In this way it is possible to see any video, audio, and transcript versions of the document and, for each version, to browse the video



The Long Island University basketball team, made up largely of ...

Fig. 2. Screenshot of the metadata editor. Document structure (*above*) and expression editing tool (*below*)

segmentations in scenes, shots, etc. By clicking on the icon corresponding to a metadata object, it is possible to modify, in a separate window, the metadata fields of the object. The Expression Tool can be used to access the metadata relative to the video segmentation and modify it. The user can view the video, hear the audio, and read the transcript. The window also shows an overview of the video segmentation by means of three slide tools (see the bottom of the Expression Tool window), which represent the video, the audio, and the transcript (if any) of the whole expression. These slides are subdivided into partitions that represent the media segmentation. By selecting a segment, the Expression Tool shows the version corresponding to the subpart of the media (for instance, a scene or a shot).

5 System prototypes

A prototype system that provides most of the functionality described so far has been developed and installed at the content providers' sites. A final prototype (ECHO Digital Film Library) was released in April 2003. It included the following enhanced functionality:

- Face detection providing support for similarity retrieval of faces,
- Detection and recognition of specific objects (e.g., cars, planes, people, etc.),
- Image analysis and similarity retrieval,
- Multimodal retrieval (based on image content and text) in order to experiment with the formulation and execution of queries combining different types of predicates,
- Support of cross-language retrieval on the audio transcript,
- Text detection and recognition.

6 Field trials and evaluation

The ECHO prototype system was evaluated in field trials conducted by two groups of end users in the education and entertainment domains. The reference standard used was ISO/IEC 9126, which defines six features to be measured: functionality, reliability, usability, efficiency, maintainability, and portability. During the project lifetime, three prototypes were delivered. The second and third prototypes were subjected to extensive user testing. The results provided important feedback for adjustments and future developments of the system. The main focus of the evaluation was on usability of the system. The usability was examined from three different perspectives:

- **Effectiveness:** the accuracy and completeness with which users achieve specific goals,
- **Efficiency:** the resources expended in relation to the accuracy and completeness of goal achievement,
- **Satisfaction:** the ease and acceptability of use.

The evaluation methodology made use of interviews, questionnaires, and observations of system usage by different user classes. Tests were made following strict predefined scenarios. All the information gathered was analyzed in order to prepare a usability report.

Overall, the results were more than satisfactory. For most of the users tested, it was the first time they had had the chance to work with a digital asset management system. Thus even the standard search option and display of results were greeted very positively, and they appreciated the possibility to browse over different collections. The more advanced functionality such as the ability to browse through transcripts and launch content-based image queries and the possibility of multilingual access was particularly welcome. However, many users did not see the need for automatically generated visual
]

summaries when they can directly access both key frames and complete documents. A number of users felt that the functionality of the system needed to be extended to support the clearing of copyright information and the billing and accounting processes. Overall, users in both application domains saw ways in which the ECHO system could be a valuable asset to their work by allowing them to access cultural heritage in a flexible way and to reuse previously inaccessible audiovisual sources.

References

1. Amato G, Castelli D, Pisani S (2000) A metadata model for historical documentary films. In: Proceedings of ECDL 2000, Lisbon, September 2000
2. ECHO (2000) http://pc-erato2.iei.pi.cnr.it/echo/
3. Fuhrt B, Smoliar SW, Zang H (1996) Video and image processing in multimedia systems. Kluwer, Dordrecht
4. Saur Muenchen KG (1998) Functional requirements for bibliographic records, final report. Available at: http://www.ifla.org/VII/s13/frbr/frbr.pdf