A Generic Annotation Model for Video Databases

Herwig Rehatschek
Institute of Information Systems, JOANNEUM RESEARCH, Graz
Steyrergasse 17 A-8010 Graz, Austria
Voice: ++43 316 876 1194 Fax: ++43 316 876 1191
E-mail: Herwig.Rehatschek@joanneum.ac.at
WWW: http://www.joanneum.ac.at/iis

Heimo Müller
Vrije Universiteit Amsterdam,
Fac. of Arts, Word & Image studies
de Boelelaan 1105, 1081 HV Amsterdam
E-mail: mue@smr.nl

Abstract: The change from analogue broadcasting to digital MPEG-2 channels among the satellite programs resulted in new demands on video databases and archives. Digital archives offer on the one hand a reduction of storage costs, and enable on the other hand easy reuse of already existing material. However, searching for appropriate film material in large archives is still a tedious problem. This paper describes a generic annotation model for MPEG movies which enables the user to structure the film in as many hierarchical levels as needed and to annotate any physical or logical part of the film with generic definable attributes. The model was implemented in the Digital Film Center system which additionally offers a query and ordering facility per web browser and Internet.

1 Introduction

An increasing number of satellites offering digital MPEG-2 channels (e.g. DF 1, Astra Service, Premiere Digital, RAI, Intelsat, ...) mark the start of a new age in the distribution of films and videos. This results in an increasing demand on content annotation in order to reuse already existing archive material for cost effective productions. However, searching for appropriate film material in a large film archive is still a tedious task. Parts of films can just be searched and retrieved if annotations are available. In praxis there are many different ways of annotation depending on the overall approach (annotation based on a thesaurus, keywords or only free text) and the application domain (broadcast archive, industrial archive, cultural archive). An additional problem occurs by using different annotation languages and country specific character sets.

When film archives are opened for commercialization or for the public the awkward handling of analogue film material becomes a problem. Digitization offers a number of advantages including reduction of storage costs, no progressive decay, fast availability in different qualities (MPEG-1/for previewing purposes, MPEG-2/for sending, ...), reusing and copying of material without loss of quality and fast access for internal personnel (Intranet) and customers (Internet).

Within the DFC (Digital Film Center) some of these problems are addressed and solved. A major focus was given on the interoperability across different application domains and the problem of

import/conversion of existing annotation data. The cross platform exchange of annotation records was studied in detail. The DFC offers two annotation possibilities, a thesaurus based and one with generic keywords in combination with free text.

2 Related Work

Several efforts are undertaken in order to define appropriate data models for storing multimedia data. One model for storing a physical, time based representation of digital video and audio was introduced by [Breiteneder, 92]. General concepts for the physical modeling of digital video and audio data are discussed and a specific model for storing Quicktime movies is introduced. The application of the general concepts allows the specific physical modeling of any other video format. The Layered Multimedia Data Model (LMDM) developed by [Schloss, 94] emphasizes the sharing of data components by dividing the process of multimedia application development into smaller pieces. LMDM claims for the separation of data, manipulation and presentation. Four layers are introduced which abstract the raw multimedia from the multimedia users and applications. LMDM defines a platform independent object oriented approach how to model multimedia data in order to guarantee later reusability on any platform. Both modeling approaches do not concentrate an the topic of generic film annotation using user definable attributes and values which can be attached to any physical or logical unit (e.g. an act, scene, shot) of a film.

A lot of research has been done on the development of digital video databases and archives. Siemens has implemented the CARAT-ARC system [Depommier, 97], which is an open system for storing, indexing and searching multimedia data. Annotation of data is supported by either using a thesaurus or free text. However, the system is not designed for supporting off-line units, e.g. outsourcing of annotation and/or encoding to geographically dispersed locations. The VideoSTAR experimental database system [Hjelsvold, 95], which was developed by the Norwegian Institute of Technology, supports storage of media files, virtual documents, video structures and video annotations in four repositories. Content based querying and retrieval of film parts is achieved by annotation of logical parts of a film (sequence, scene, shot, compound units). Despite the relational data model of VideoSTAR offers annotation it is not generic in the sense that users can define new categories but limited to four categories, which can hold free text.

There exist several sites offering search for film meta-information and download of movie clips on the internet. Some just offer an alphabetically ordered list of films with previews, others offer a database system with access to stored film meta-information. The Internet Movie Database (IMDb) [IMDb, 98] covered in January 1998 approximately 120,000 movies with over 1,750,000 filmography entries. The database aims to capture information associated with movies from across the world, starting with the earliest cinema, going through to the very latest releases and even movies still in production. The database can be queried for various of meta-information. However, the IMDb treats films as a unit, so it does not support a search for scenes, acts and shots of a film. Furthermore IMDb does not offer movie previews and direct ordering. Copyright issues are not covered within the database.

Film.com provides a collection of material on film for a world-wide audience and bringing together critics, writers, and movie buffs [FilmCom, 98]. Film.com offers a number of services including: a store, official film reviews, a discussion forum and a small number film previews in three categories via RealVideo. All film previews are offered in a alphabetically ordered list, no search facility for film meta-information (e.g. physical storage media, producer, director, actors, ...) is available. The on-line shop is separated into the categories merchandise (posters, T-shirts, etc.), music, books, gifts and videos. All entities in the shop are indexed by title and cannot be searched. Films are not annotated, therefore parts of films can neither be searched nor ordered.

3 The Digital Film Center – DFC

This section gives an overview of the DFC system architecture and its high level building blocks. The DFC system is a very large digital video database holding all films in MPEG-2 format. Sources remain stored on Digital Betacam in order to fulfill any special format wishes of customers (e.g. S-VHS). Each film has annotations attached which allows the search of specific parts or objects (e.g.: acts, scenes, shots, actors, ...) in a film. Basically the DFC system consists of four units (see Figure 1): compression street(s), annotation site(s), central DFC digital video database and the web interface for on-line search and ordering.

According to Figure 1 the filling process of the database can be described as follows: incoming videos are first encoded at the compression sites in two formats: MPEG-2 for storage in the central DFC video tape archive and resell and MPEG-1 for low resolution previews and annotation purposes. The encoded material is then sent together with some film metainformation (production year, producer, director, length, time codes, MPEG parameter, and so on) to the central DFC video database on DLT tapes. The metainformation is stored in SGML [ISO, 86] format in order to make the system as open as possible. The metainformation of the film is imported, the MPEG data remains on the DLT tape. The database stores a reference to the tape and the location for later access. Now the film is ready for annotation and can be checked out by an annotation site. For this purpose the MPEG-1 representation together with the already existing film metainformation is sent to an annotation site using again SGML as an exchange format. At the annotation site the film is being annotated using a special annotation software. The annotation is sent back in the same SGML based format to the central DFC database. Now information about the film and all parts of the film is contained in the video database. These information can be searched by customers via a web interface. Because of the attached annotations the search of parts of the film or specific objects within the film become possible. Any parts of the films can later be ordered on-line vie the web interface.

Since the compression streets and the annotation sites have a special SGML based off-line interface for importing / exporting information and data to the central DFC video database these units can be built at geographically dispersed locations all over the world. Encoded MPEG data is sent via DLT tapes (which also contain some metainformation of the encoded film), annotations can be sent via any exchangeable media (e.g. CD-ROM, ZIP, ...) to the central DFC video database.

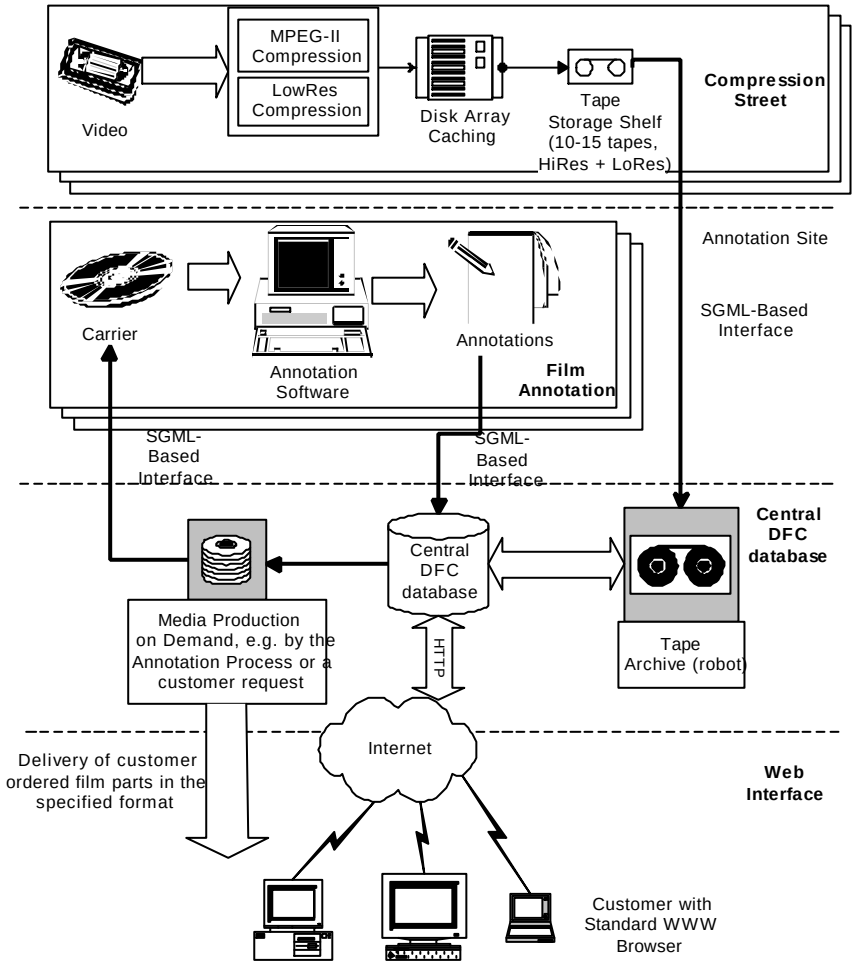

Figure 1: High level building blocks of the Digital Film Center

4 The Generic Annotation Model

The most important and central component of the DFC system is the digital video database. It holds all information related to films and parts of films. Within the database there exist two main views on films: the logical and the physical view. The starting point is the logical film. It has several physical representations, and is the target of annotations. This is different to current systems, where in most cases a physical representation of a film is annotated. Both views are modeled in the DFC data scheme.

One physical representation is the reference source. When describing differences in terms of annotation of different representations (e.g. different language versions, or evening versus late night versions) all annotations are made relative (or in reference) to the time codes of the reference version.

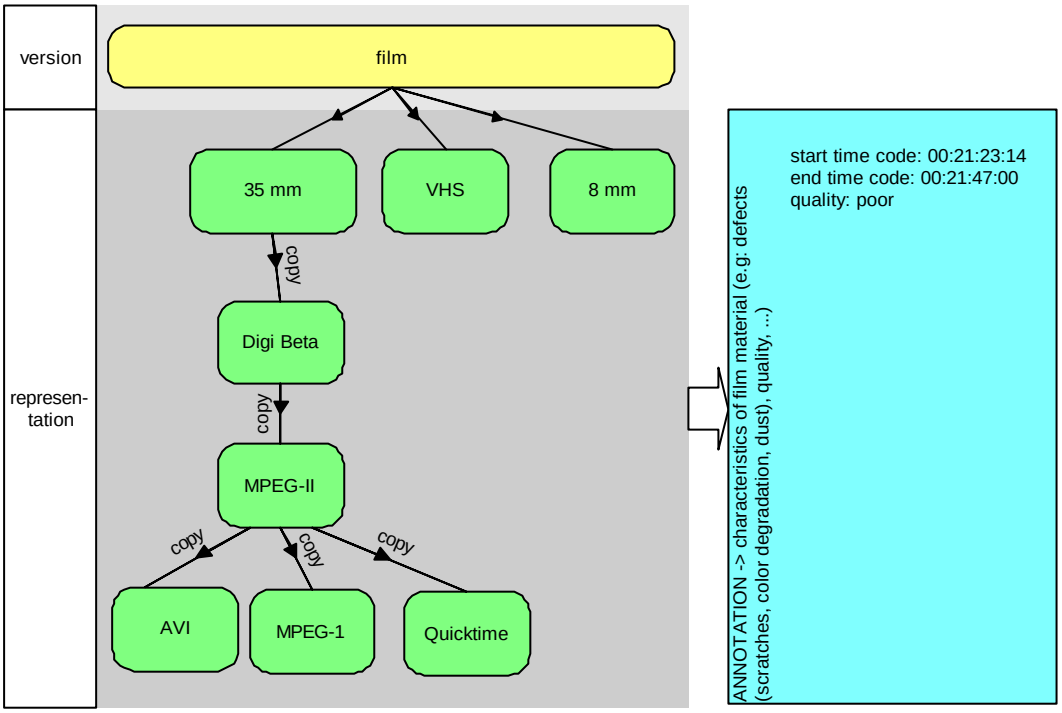

Figure 2: Physical representation of an example film

We want to stress the fact, that when annotating a film there exist basic semantics: the temporal static structure of the film (referred to as static annotation) and annotations, which vary in their semantics (referred to as dynamic annotation). E.g., when annotating a video database of a hospital, person annotation will describe patients, when describing news material, we annotate real world persons and their historic actions, and in the annotation of a movie, actors/characters of the movie are described. The annotation model of the DFC database, therefore defines a model to describe the basic semantics (temporal and logical structure of video) and provides a method to describe the dynamic part of the annotation data.

The temporal model allows to construct any given structure of a film in as many levels as needed. Subdivisions of films into several units are supported (e.g. a film may be divided into several acts; an act may consist of several scenes and a scene may be divided into several shots where each shot has a special frame of interest). The data model consists of the units parts, sequences and groups. Parts are the smallest units. They address a number of frames (or even just one) and are defined by a start and end time code. Sequences can be defined recursive and can therefore again contain sequences. This allows a modelling of as many levels as needed. Besides sequences groups can be formed which represent any combination of parts, sequences and also again groups. Groups do not have the requirement to contain continuous sequences of time codes and are therefore a good instrument to structure a film according to key scenes (e.g. for a trailer or advertisement production). Since a film can have more than one version these structures can exist for each version and can be actually different. The DFC video database supports versions and can hold a structure for each of them. This is indicated by the version level at the bottom of Figure 3.

An example for the temporal structuring of a movie is given in Figure 3. The film "Opernball" is structured in acts, scenes and shots using our hierarchical film data model. In this example parts represent shots, sequences of first order scenes and sequences of second order acts.

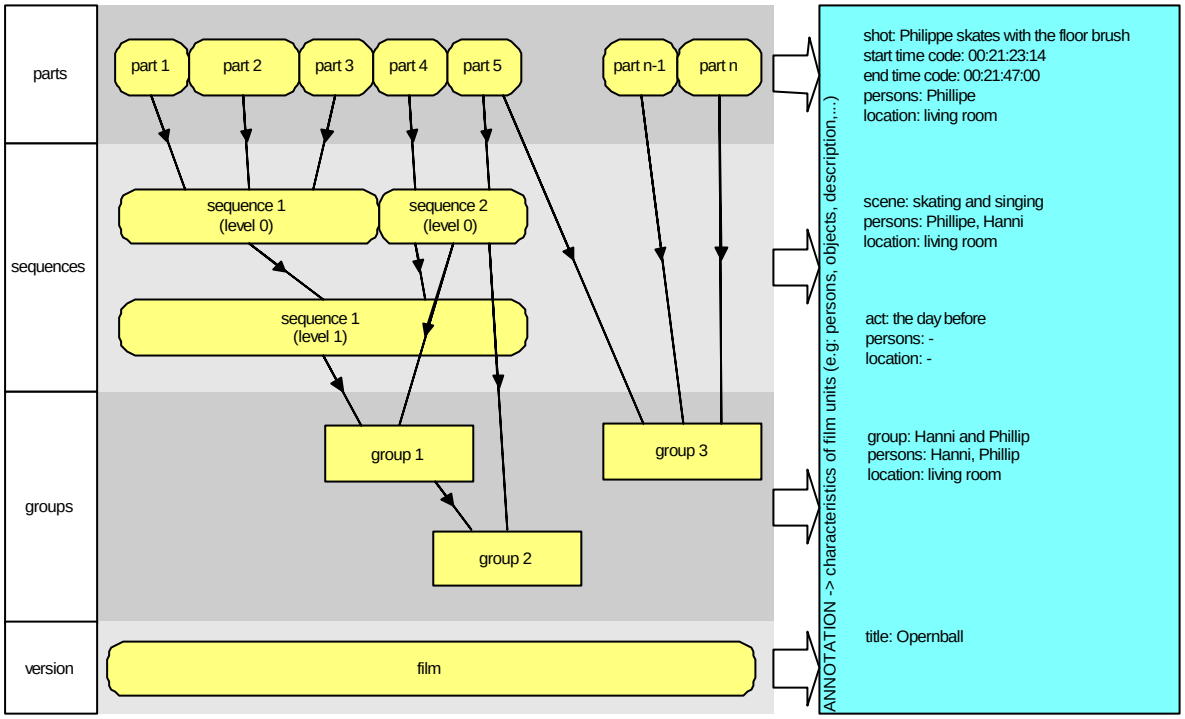

Figure 3: Logical structure of an example film

All entities of the logical and physical structure of a film can be annotated and therefore also be searched. The semantics of such an annotation is defined in the so called "annotation style file". An annotation style file holds a number of annotation attributes. Annotation attributes can be defined generic, in the sense that the user can define the attribute's name and its type. One annotation style concentrates on one special kind of movie, e.g. a medical film or a documentary film, and has therefore special annotation attributes. E.g. for a documentary film some attributes could be "geography / city" (type text), "animal / mammal / body temperature" (type number). Different styles can be created and stored in the database. Annotation styles are defined by using SGML [ISO, 86] in combination with natural language description. The set of all annotation style used is called video object foundation class. In Figure 4 the basic idea of generic film annotation is summarized.

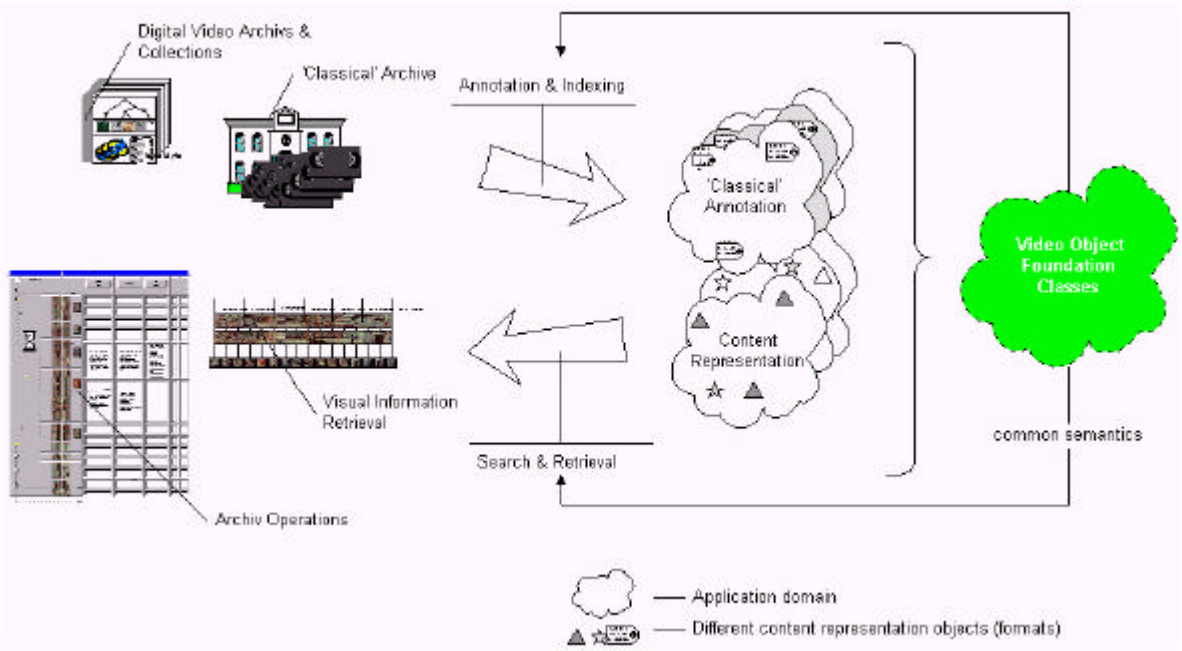

Figure 4: The basic idea of generic film annotation

The generic implementation within the RDBMS was implemented by defining attributes by a name, a detailed description, a data type and a possible default value. Such defined attributes can now be assigned to any logical or physical film entities (e.g. parts, sequences, physical representations, etc.). Two examples for generic annotation attributes having the type "text" and "date" are as follows:

name:	objects
description:	objects that can be seen on a film sequence
data type:	text
default value:	empty
comment:	This text attribute will be assigned to the entities acts, scenes, shots. After the assignment objects are part of the annotation for acts, scenes and shots.

name:	production date
description:	production year of a film (e.g.: 1994)
data type:	date
default date:	empty
comment:	This date attribute is assigned to the entity film. After the assignment of this attribute the production year of a film may be set by an annotator.

Next to the generic annotation DFC supports a thesaurus based keyword annotation. The thesaurus can be defined by the user with a special tool and stored within the DFC database. The database supports different thesauri according to different kind of movies.

Based on this generic annotation model the annotation process takes place. With a special piece of software the generation of the described logical and physical structures is done via a graphical user interface. For the annotation process low resolution video data (MPEG-I), already existing annotation data and basic film description data is transported on carriers (e.g. CD-ROM, Optical, ...) to an annotation site. Annotation is done locally and the results are returned to the central DFC database.

The graphical user interface of the annotation software allows generic annotation in four user definable levels, where each level can be annotated separately. For example, speaking in terms of film, these four levels could refer to acts, scenes, shots and frames. The GUI also visualizes key frames which help the user to differentiate among different defined annotation parts (e.g. scene 1 is characterized by key frame 1). Static annotation allows to fill in all information concerning the whole film, e.g. producer, director, etc. Generic annotation is done by keywords according to the annotation styles which are defined in the database.

All annotation and encoded data will be stored in the central DFC database which is accessible over the world-wide web to customers for searching and ordering. The web interface provides access to the database material for customers. By filling an electronic shopping cart authorized customers can order the desired film material - which actually can be parts of a film - in the desired quality. The querying possibilities offered support the search for generic annotation attributes as well as free text search. In Figure 5 one typical search result is given. The result represents one of the parts of the film "Opernball" which have been returned on the query "return all parts which contain a table in any annotation attribute". Next to the detailed description of the part including start and end time code and all the generic annotation attributes authorized customers are able to preview material by clicking on a link within the result. Previews are stored on hard disks for fast access and not on mass storage devices, where the high-quality material of the archive is kept.

5 Results and Conclusions

This paper addressed a digital video database system, referred to "DFC - Digital Film Center", which allows (1) storage of digital videos and corresponding metainformation (2) generic annotation of user defined film structures (3) search access on annotation data via a web interface and a standard WWW browser.

The DFC video database system is designed as a large geographically dispersed system. Many encoding suites produce MPEG-2 videos, one central video database holds metainformation, annotations and references to the MPEG-2 files stored in a tape robot, many annotation sites add film annotations to the stored movies. The central database can be accessed via a web interface by a standard WWW browser all over the world.

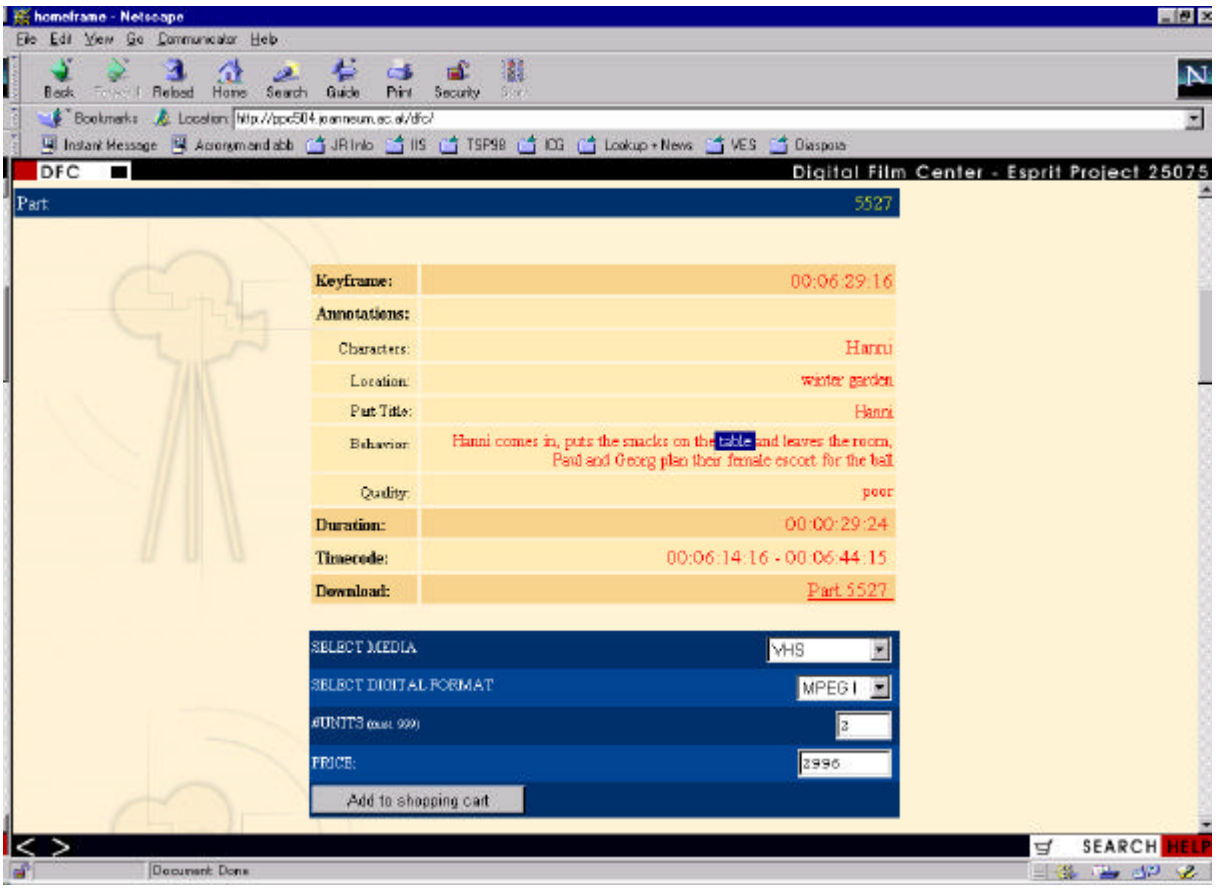

Figure 5: Web interface

The generic film data model of the DFC system allows the hierarchical structuring of a film in as many levels as needed. This can be done on the one hand for the logical structure (e.g. acts, scenes, shots and frames) and on the other hand for the physical representation of a film. To each of these logical and physical entities annotations can be attached.

The generic annotation model is the most remarkable part of the DFC video database. The generic annotation model allows the free definition of annotation attributes with any user defined name and type. These annotation attributes can be structured in so called "annotation styles". Different annotation styles can be stored in the video database. One style refers to one specific annotation topic (e.g. medical films, action films, ...). The generic annotation is done by a special annotation software which supports the annotator with a graphical user interface and a MPEG-I preview. A second annotation possibility is thesaurus-keyword based, where the thesaurus can be dynamically created and exchanged.

A web interface was developed in order to search the database and download previews. The web interface offers registered users the search for entire films (e.g. title search) and parts of a film. Search results can be collected in a shopping cart and on-line ordering can take place. The quality of the ordered film material can be chosen by the customer.

The DFC system does not use a proprietary exchange format among the distributed units. All interfaces between the central video database, the annotation software and the encoding suites are SGML-based which makes the DFC system an open system. Imports from and exports to other video database systems, e.g. Media Vault, become possible.

6 Outlook

Currently annotation styles are defined with SGML and natural language description. In the future formal specification methods could be used for describing the semantics of the annotation fields, and their relations.

The development of Video Object Foundation Classes will be stressed in the future, which describe a framework of basic objects semantics, e.g. persons, settings, speech, movement patterns, and methods of specializing these objects for a specific annotation style.

The new member of the MPEG family, called "Multimedia Content Description Interface" (in short ‘MPEG-7’), will extend the limited capabilities of proprietary solutions in identifying existing content notably by specifying a standard set of descriptors that can be used to describe various types of multimedia information. Developments on this standard will be closely monitored and checked for integration into the DFC system.

7 Acknowledgments

The Digital Film Center was an ESPRIT project (Nr. 25075) and was partially funded by the European Union. Specific thanks go to our colleagues Bernd Jandl and Harald Mayer and to our Greek partners within the project which helped to realize this system.

8 References

[Arman, 94] F. Arman, R. Depommier, A. Hsu, M. Chiu, R. Depommier. “Content-Based Browsing of Video Sequences“, ACM Multimedia, 1994

[Breiteneder, 92] Ch. Breiteneder, S. Gibbs, D. Tsichritzis. "Modelling of Audio/Video Data", pp. 322-339, Karlsruhe, Germany, 1992

[Depommier, 97] R. Depommier, N. Fan, K. Gunaseelan, R. Hjelsvold. "CARAT-ARC: A scalable and Reliable Digital Media Archiving System", IEEE International Conference on Image Processing, 1997

[England, 96] P. England, R. Allen, M. Sullivan, A. Heybey, M. Bianchi, A. Dailianas. "The Bellcore Video Library Toolkit. Storage and Retrieval for Image and Video Databases“, pp. 254-264, San Diego/La Jolla, CA, USA 1996

[FilmCom, 98] Film.com Inc., January 1998. <URL: http://www.film.com/admin/about.htm>

[Hjelsvold, 95] R. Hjelsvold, R. Midtstraum. "Databases for Video Information Sharing", Proceedings of the IS&T/SPIE Symposium on Electronic Imaging Science and Technology, Conference on Storage and Retrieval for Image and Video Databases III, San Jose, C.A., February 1995.

[ISO, 86] International Organization for Standardization, 1986. *Information processing---Text and office systems---Standard Generalized Markup Language (SGML)*, Geneva ISO, 1986.

[IMDb, 98] The Internet Movie Database, January 1998. <URL: http://us.imdb.com>

[Schloss, 94] Schloss, Wynblatt. “ Providing Definition and Temporal Structure for MM data“ , Proceedings of the Second ACM International Conference on Multimedia, CA., ACM Press, ISBN 0-89791-686-7, San Francisco, 1994