Intelligent Semantic Access to Audiovisual Content

Yannis Avrithis[1], Giorgos Stamou[1], Anastasios Delopoulos[2], and Stefanos Kollias[1]

[1] Image, Video and Multimedia Systems Laboratory
Department of Electrical and Computer Engineering
National Technical University of Athens
9, Iroon Polytechniou Str., 15773 Zographou, Athens, Greece
{iavr,gstam}@image.ntua.gr, stefanos@cs.ntua.gr

[2] Division of Electronics & Computer Engineering
Department of Electrical and Computer Engineering
Faculty of Engineering, Aristotle University of Thessaloniki
Thessaloniki 54006, Greece
adelo@eng.auth.gr

Abstract. In this paper, an integrated information system is presented that offers enhanced search and retrieval capabilities to users of heterogeneous digital audiovisual (a/v) archives. This novel system exploits the advances in handling a/v content and related metadata, as introduced by MPEG-4 and worked out by MPEG-7, to offer advanced access services characterized by the tri-fold "semantic phrasing of the request (query)", "unified handling" and "personalized response". The proposed system is targeting the intelligent extraction of semantic information from a/v and text related data taking into account the nature of useful queries that users may issue, and the context determined by user profiles. From a technical point of view, it will play the role of an intermediate access server residing between the end users and multiple heterogeneous audiovisual archives organized according to new MPEG standards.

1 Introduction

Digital archiving of multimedia content including video, audio, still images and various types of documents has been recognized by content holding organizations as a mature choice for the preservation, preview and partial distribution of their assets. The advances in computer and data networks along with the success of standardization efforts of MPEG and JPEG boosted the movement of the archives towards the conversion of their fragile and manually indexed material to digital, computer accessible data. By the end of last century the question was not on whether digital archives are technically and economically viable, but rather on how digital archives would be *efficient* and *informative*. In this framework, different scientific fields such as, on the one hand, development of database management systems, and on the other hand, processing and analysis of multimedia data, as well as artificial and computational intelligence methods, have observed a close cooperation with each other during the last few years. The attempt has been to develop intelligent and efficient human computer inter-

I.P. Vlahavas and C.D. Spyropoulos (Eds.): SETN 2002, LNAI 2308, pp. 215–224, 2002.

action systems, enabling the user to access vast amounts of heterogeneous information, stored in different sites and archives.

Database management systems (DBMS) have been designed that are able to handle such types of access to the stored information. Attaching information bits, called metadata, to the original data is the means for achieving this goal. The focus of technological attempts has been on the analysis of digital video, due to its large amounts of spatio-temporal interrelations, which turns it into the most demanding and complex data structure. Current and evolving international standardization activities, such as of the EBU, MPEG-4 [3,4,9], MPEG-7 [5-8], or JPEG-2000 [10] for still images, deal with aspects related to data structures and metadata. In particular, the new MPEG standards are object-oriented, i.e., adopt video objects as the information units, which is different from the information units used in the current form of video and film, i.e. scenes or shots. Of major importance is the contribution of MPEG-7 and JPEG-2000 to using metadata related to the visual and acoustic content of archived objects.

In more detail, MPEG-7 will define a standard for describing multimedia content. The objective is to quickly and efficiently search and retrieve audiovisual material. To allow interoperability, the standard adopts some normative elements, such as Descriptors (D’s), Description Schemes (DS’s), the Description Definition Language (DDL) as well as Coding and System Tools. The Descriptors define the syntax and the semantics of the representation of features, while the Description Schemes specify the structure and semantics of the relationships between Descriptors or other Descriptions. Many descriptors have been submitted for MPEG-7, some of which either accepted and included in the eXperimental Model (XM), which is a platform and tool set to evaluate and improve the tools of MPEG-7, or are in the experimentation (Core Experiments, CE) phase. Two parallel levels of descriptors are defined: the syntactic one, which describes the perceptual properties of the content, such as color and motion of spatio-temporal segments and the semantic one, which describes the meaning of content, in terms of semantic objects and events. Syntactic description seems to be well in hand in MPEG-7, but fleshing out the semantic description has not yet received the required attention.

It becomes clear among the research community dealing with content-based audiovisual data retrieval and new emerging related standards such as MPEG-7, that the results to be obtained will be ineffective, unless major focus is given to the semantic information level, defining what most users desire to retrieve. Mapping, however, low level, subsymbolic descriptors of a/v archives to high level symbolic ones is in general difficult, even impossible with the current state of technology. It can, however, be tackled when dealing with specific application domains. It seems that the extraction of semantic information from a/v and text related data is tractable taking into account [1]:

- *The nature of useful queries that users may issue*. This is only a portion of the general set of questions related to "content understanding". Using all types of multimedia information of the archives makes the task more tractable.
- *The context determined by user profiles.*

In this paper a novel platform is proposed that intends to exploit the aforementioned ideas in order to offer user friendly, highly informative access to distributed audiovisual archives.

2 Architecture of the Proposed System

The general architecture is provided in Figure 1, where all modules and subsystems are depicted, but the flow of information between modules is not shown for clarity. More detailed system diagrams and descriptions of subsystems are provided in the following sections for the two main modes of system operation, i.e. *update mode* and *query mode*. The system has the following features:

- Adopts the general features and descriptions for content-based access to visual information proposed by MPEG-7 and other standards such as JPEG-2000; also adopts existing basic system architectures implementing the MPEG-4 and MPEG-7 standardisation activities.
- Performs dynamic extraction of high level semantic description of a/v content units (movies, scenes, shots, etc.) on the basis of syntactic and lower level semantic information contained in the a/v archives.
- Enables the issuing of queries at a high semantic level. This feature is essential for unifying user access to multiple heterogeneous a/v archives with different structure and description detail.
- Generates, updates and manages users' profile metadata that specify their preferences against the a/v content.
- Employs the above users' metadata structures for filtering the information returned as response to their queries so that it better fits to user preferences and priorities. To this end static, adaptive or dynamic classification of the available a/v content is performed by the a/v classification module, and next "compared" to individual users' profiles.
- Gives users the ability to define and redefine their initial profile.
- Is capable to communicate with existing a/v archives, structured on the basis of scenes/shots and key frames, or with already developed systems with proprietary user interfaces. In the former case, it will permit translation of the basic information units to more complex object-based ones; in the latter, it will accept and adapt a/v data, objects and stored metadata.
- User interfaces employ platform independent tools targeting both the Internet and WWW and broadcast type of access routes.

Additionally, it is important that the system has the following features related to user query processing:

Response time: Internal intelligent modules may use semantic information available in the DBMS (calculated by *Dynamic Thematic Categorization*-DTC and *Detection of Events and Composite Objects*-DECO) to locate and rank multimedia documents very fast, and some times without querying individual a/v archives. In most cases where a/v unit descriptions are required, query processing may be slower due to the large volume of information. In all cases it is important that the overall response time of the system is not too long as perceived by the end user.

Filtering: When a user specifies a composite query, it is desirable that a semantic query interpretation is constructed and multimedia documents are filtered as much as possible according to the semantic interpretation and the user profile, in order to avoid the overwhelming responses of most search engines.

Exact matching: In the special cases where the user query is simple, e.g. a single keyword, the system must return all documents whose description contains the keyword; no information is lost this way.

Ranking: In all cases retrieved documents must be ranked according to the user's preferences and their semantic relevance to the query, so that most important documents are presented first.

Up-to-date information: Since the system is designed for handling a large number of individual a/v archives whose content may change frequently, DBMS must be updated (either in batch updates or updates on demand) to reflect the most recent archive content.

Relevance feedback: It will be useful and probably necessary to provide a relevance feedback mechanism to permit refinement of user queries. Used in modern information retrieval systems, this mechanism allows the user to select those documents among the first retrieval results that are most "relevant" to the original query; the latter is then automatically refined to retrieve similar documents.

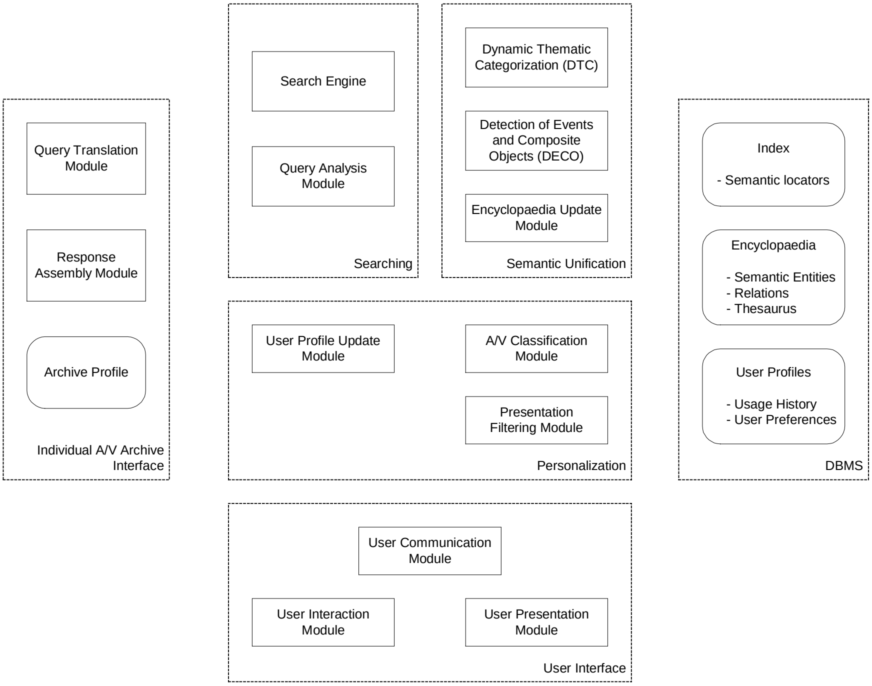

Fig. 1. General architecture of the system

The description of subsystems functionality follows the distinction in two main modes of operation. In *query mode*, the system is online and used to process user requests by translating/dispatching queries to the archives and assembling/presenting the respective responses. The main internal modules participating in this mode are *query analysis*, *search engine*, *a/v classification* and *presentation filtering*.

An additional *update mode* of operation will also be necessary for updating the content description data. In particular, a batch update procedure can be employed at regular intervals to perform DTC and DECO on available a/v units and update the database. Alternatively, an *update on demand* procedure can be employed whenever new a/v units are added to individual a/v archives to keep the system synchronised at all times. The decision will depend on speed, storage and network traffic performance considerations. The main internal modules participating in the update mode are *DTC*, *DECO*, *encyclopaedia update* and *user profile update*.

An overview of the functionality of the subsystems and modules is described below in two separate sections for the query mode and the update mode, where additional diagrams depict detailed flow of information between modules.

3 Query Mode of Operation

In query mode, the system is online accepting user requests, translating / dispatching queries to the archives and assembling / presenting the respective responses. The main internal modules participating in this mode are *query analysis*, *search engine*, *a/v classification* and *presentation filtering*. The overall diagram of this mode of operation is depicted in Figure 2.

The user query is first submitted at the *user interaction* module of the *user interface*. It mainly consists of two parts:

- *Semantic specification*: either keywords with logical operators, free text or composite, structured statements/scenaria designed by special forms and input controls. It may also contain audiovisual content features specified either through text or special input controls.
- *Metadata specifications*: keywords, numbers, dates etc. representing metadata specifications such as creation, media, usage, classification, navigation and access information as defined in MPEG-7.

The metadata part of a query will be finally dispatched to individual a/v archives (after translation at the archive interfaces); the semantic part however is first processed within the internal intelligent modules of Feathon to accommodate for semantic unification. The query is first transformed in a suitable structure (*user query*) by the communication module of the user interface and then transferred at the *query analysis* module of the *searching* subsystem. This module performs three main operations:

- *Query interpretation*: receives the semantic part of queries in the form of vectors or graphs of keywords and replaces the keywords by *semantic entities* (objects, events, concepts, agents etc.) found in the *encyclopaedia*.
- *Query expansion*: takes advantage of the semantic entity relations in the encyclopaedia and the *thesaurus* (an automatically updated association table between semantic entities) to expand queries using entities that do not appear in the original query. E.g. a goal event in football can be expanded in a suitable combination of objects such as a player, a ball and a goal post [2].

- *Profiling*: adds relevant information from the *user preferences* of the user profile (e.g. interest for European or American football) and adjusts the query accordingly to perform *pre-filtering*. The user preferences, apart from the normative elements described in MPEG-7, may contain thematic categories and interests in the form of composite semantic entities.

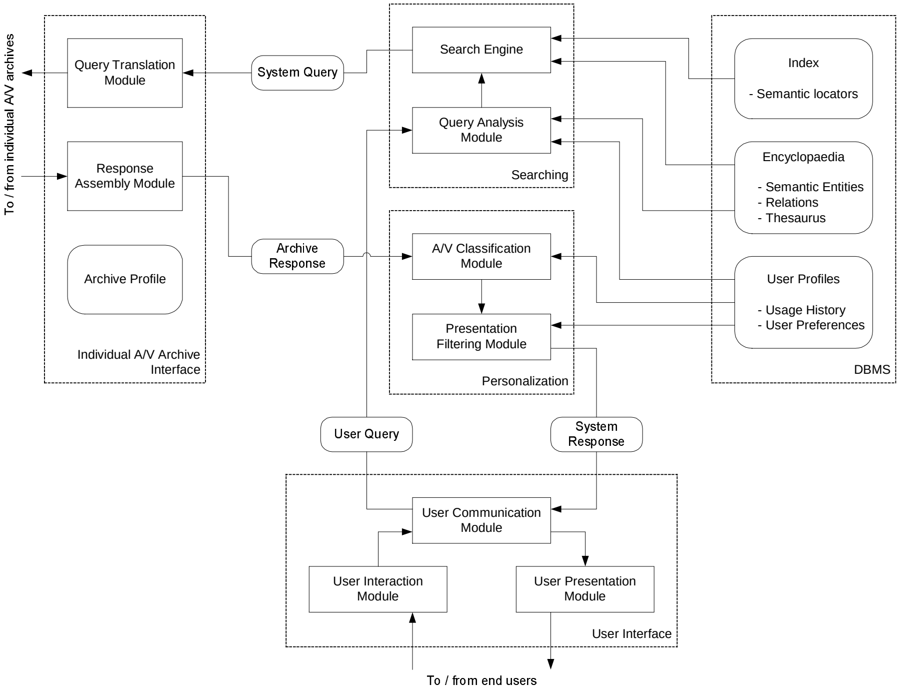

Fig. 2. The system at query mode of operation

The final result of query analysis, the *internal query*, is a structure (vector or graph) of semantic entities along with confidence values. This is transferred to the *search engine* where this structure is tested against the *index*. The index contains sets of document locators for each thematic category and semantic entity of the encyclopaedia (and also for a large set of composite entities), resulting from DTC and DECO procedures. The result is a list of document locators corresponding to documents at different a/v archives. This list is combined by the search engine with the metadata part of the user query to construct the *system query*, which is a unified query dispatched to all individual a/v archive interfaces.

At each *a/v archive interface*, the *query translation* module uses the *archive profile* to associate normative DS's of the system query to the proprietary structures employed in each archive. This translation takes place only for the metadata part of the query; the semantic part has already produced a known list of media locators. Then, depending on the archive interfacing type, the interface either dispatches the query to an existing *archive search engine* (through the query translation module) or communicates directly with the multimedia document descriptions (or even the multimedia documents themselves) included in the archive. The result is a filtered (but not ranked) list of document locators (or links) and possibly their descriptions. The *response assembly* module of the interface constructs the *archive response* as a unified structure of retrieved documents, which is returned to the a/v classification module of the *personalisation* subsystem.

The *a/v classification module* performs ranking (but not filtering) to the retrieved documents of the archive response based on *user interests* contained within the preferences of the user profiles. The user interests consist of (unified) thematic categories and simple or composite semantic entities of the encyclopaedia. Dynamic categorisation and detection of composite entities is performed on the retrieved documents using their entire descriptions and relevance values are assigned after matching with the user interests. The result is the *internal response*, which is a ranked list of a/v documents with their descriptions.

The internal response is transferred to the *presentation filtering* module where further ranking and filtering is performed according to the remaining parts of the user preferences such as creation, media, classification, usage, access, and navigation preferences (e.g. favourite actors / directors or preference for short summaries). The *system response* produced by the presentation filtering module, which is also a ranked list of a/v documents with their descriptions, is transferred to the *communications module* of the user interface and finally to the *user presentation* module. The entire record of user actions during the search procedure (user query, retrieved documents, documents selected as relevant) is stored in the *usage history* of the specific user; this information is then used for tracking and updating the user preferences. Furthermore, *relevance feedback* is supported by the system. This would require modifications in the user interaction and presentation modules of the user interface, and a feedback module in the personalisation subsystem.

4 Update Mode of Operation

The general scope of the information update mode of operation is to adapt and enrich the DBMS used for the unified searching and filtering of a/v content. Its operation is based on the *semantic unification* and the *personalisation* subsystems depicted in Figure 3. The semantic unification subsystem is responsible for the construction and update of the *index* and the *encyclopaedia*, while the personalisation subsystem updates the *user profiles*.

As already mentioned, the index stored in the DBMS consists of a set of semantic document locators, i.e. a set of encyclopaedia terms with links to the a/v unit descriptions (stored in the a/v archives) that semantically “contain” them. The information units of the index are semantic entities (objects, events, concepts, thematic categories, etc.) stored in the encyclopaedia, and composite semantic structures (relations, composite objects or scenarios, probably not contained in the encyclopaedia) representing the abstract semantic meaning of complex concepts and events. It is mentioned that although thematic categories are actually a special case of concepts, they are stored and processed as separate units due to their important role during the searching process.

The modules that update the terms of the index and its links to the a/v units are DTC and DECO. The former takes the thematic categories stored in the encyclopaedia as input, unifies them with the thematic categories of the a/v archives and scans the a/v units in order to find and store (as links) the a/v units that belong to each thematic category, together with a weight representing the degree in which the system believes that the a/v unit is characterised of this thematic category. The latter performs a similar task for the objects, events and concepts of the encyclopaedia. Furthermore, it scans the a/v units and searches for composite semantic structures and links them with the corresponding a/v units.

All update procedures may be performed globally for the entire content of the a/v archives at regular intervals (*batch update*) or whenever the a/v content of an archive is updates (*update on demand*). In the latter case, which is preferable due to low computational cost, the update process is *incremental*, i.e. only the newly inserted a/v unit descriptions are necessary.

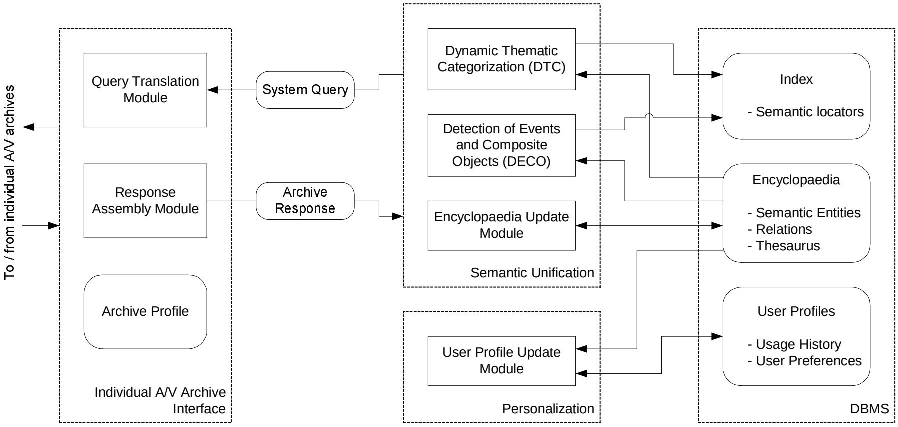

Fig. 3. The system at update mode of operation

The content of the encyclopaedia is updated with the aid of the *encyclopaedia update* module. The main goal of this module is to update the thesaurus that associates semantic entities through semantic relations. Moreover, the semantic entities of the encyclopaedia should be updated, especially when the content of the a/v archives is dramatically changed. Finally, new terms may be inserted in the encyclopaedia (especially composite semantic structures) after the mining process of the DECO and their insertion in the index.

One of the most important tasks of the update mode of operation is the update of the user profiles. This is carried out with the aid of the *user profile update* module of the *personalisation* subsystem. The structures of the DBMS that should be updated within this process are the usage history and the user preferences. The *usage history* is be updated after the end of a user query by storing all transactions of the user during the query process. The above transactions characterise the user and express his personal view of the a/v content. The user profile update module takes these transactions as input and with the aid of the Encyclopaedia and the multimedia descriptions of the a/v units referred to in the usage history, extracts the *user preferences* and stores them in the corresponding user profile. The user preferences are actually a set of semantic entities (objects, events, concepts, thematic categories) taken from the encyclopaedia, with the corresponding weights. Furthermore, they contain a set of more abstract semantic concepts (not as general as the thematic categories), the *interests*. The interests are extracted from the multimedia descriptions of the a/v units selected by the user, through a data mining process.

5 Conclusions

The core technological target of the system is to blend the achievements in characterizing a/v content - especially visual and acoustical content - with state of the art hybrid intelligence technologies in order to

(i) offer unified semantic views to existing a/v archives, if possible, beyond the individual classification schemes and subject indexes of each archive

(ii) personalize those views according to the retained profile of individual users or specific user groups; the latter clearly appreciating that semantic interpretation heavily relies on the context which in turn depends on the specific profile.

The system provides novel tools and methods for extracting high-level semantic information. Finally, using statistical and relevance feedback techniques are used to assist personalization.

Acknowledgements. This work has been partially funded by the project FAETHON of the Information Society Technologies (IST) programme of the European Community.

References

1. Delopoulos, A., Kollias, S., Avrithis, Y., Haas, W., Majcen, K.: Unified Intelligent Access to Heterogeneous Audiovisual Content. In Proc. of Int. Workshop on Content-Based Multimedia Indexing (CBMI), Brescia, Italy, Sept. 2001
2. Akrivas, G., Stamou, G., Kollias, S.: Fuzzy Semantic Association of Audiovisual Document Descriptions. In Proc. of Int. Workshop on Very Low Bitrate Video Coding (VLBV), Athens, Greece, Oct. 2001
3. Battista, S., Casalino, F., Lande C.: MPEG-4: A Multimedia Standard for the Third Millenium, Part 1. IEEE Multimedia **6** (4) (1999) 74-83

4. Battista, S., Casalino, F., Lande, C.: MPEG-4: A Multimedia Standard for the Third Millenium, Part 2. IEEE Multimedia **7** (1) (2000) 76-84
5. Nack, F., Lindsay, A.: Everything You Wanted to Know About MPEG-7: Part 1. IEEE Multimedia **6** (3) (1999) 65-77
6. Nack, F., Lindsay, A.: Everything You Wanted to Know About MPEG-7: Part 2. IEEE Multimedia **6** (4) (1999) 64-73
7. Special Issue on MPEG-7. IEEE Trans. On Circuits and Systems for Video Technology **11** (6) (2001) 685-772
8. ISO/IEC JTC1/SC29/WG11 N4032: Introduction to MPEG-7. Singapure (2001)
9. ISO/IEC JTC1/SC29/WG11 N3747: MPEG-4 Overview (V.16 – La Baule Version). La Baule, France (2000)
10. ISO/IEC JTC1/SC29/WG1 N1646R: JPEG 2000 Part I Final Committee Draft Version 1.0 (2000)