Adding Semantics to Audiovisual Content: The FAETHON Project

Thanos Athanasiadis and Yannis Avrithis

Image, Video and Multimedia Systems Laboratory
School of Electrical and Computer Engineering
National Technical University of Athens
9, Iroon Polytechniou St., 157 73 Zographou, Greece
{thanos, iavr}@image.ntua.gr

Abstract. This paper presents FAETHON, a distributed information system that offers enhanced search and retrieval capabilities to users interacting with digital audiovisual (a/v) archives. Its novelty primarily originates in the unified intelligent access to heterogeneous a/v content. The paper emphasizes the features that provide enhanced search and retrieval capabilities to users, as well as intelligent management of the a/v content by content creators / distributors. It describes the system's main components, the intelligent metadata creation package, the a/v search engine & portal, and the MPEG-7 compliant a/v archive interfaces. Finally, it provides ideas on the positioning of FAETHON in the market of a/v archives and video indexing and retrieval.

1 Introduction

In less than ten years the World Wide Web has evolved into a vast information, communication and transaction space. Needless to say its features differ greatly from those of traditional media. Projects and related activities supported under the R&D programs of the European Commission have made significant contributions to developing:

- New models, methods, technologies and systems for creating, processing, managing, networking, accessing and exploiting digital content, including audiovisual (a/v) content.
- New technological and business models for representing information, knowledge and know-how.
- Applications-oriented research – focusing on publishing, audiovisual, culture, education and training – as well as generic research in language and content technologies for all applications areas.

In this framework, the IST project FAETHON [1], has been an approach towards realising the full potential of globally distributed systems that achieve information access and use. Of primary importance is FAETHON's contribution towards the Semantic Web [2]. The fundamental prerequisite of the Semantic Web is "making content machine-understandable"; this happens when content is bound to some formal description of itself, usually referred to as "metadata". Adding "semantics to content"

in the framework of FAETHON is achieved through algorithmic, intelligent content analysis and learning processes.

FAETHON has closely followed the developments of MPEG-7 [3] and MPEG-21 [4] standardization activities and successfully convolved technologies in the fields of computational intelligence, statistics, database technology, image/video processing, audiovisual descriptions and user interfaces to build, validate and demonstrate a novel intermediate agent between users and audiovisual archives. The overall objective of FAETHON has been to develop a stand-alone, distributed information system that offers enhanced search and retrieval capabilities to users interacting with digital audiovisual archives [5]. The project outcome contributes towards making access to multimedia information, which is met in all aspects of everyday life, more effective and more efficient by providing a user-friendly environment.

This paper is organized as follows: Section 2 presents an overview of the FAETHON system. Section 3 describes the FAETHON intelligent metadata creation package, focusing on the system's offline content processing capabilities. Section 4 deals with the FAETHON a/v search engine and portal, focusing on its advanced semantic search and personalization functionalities. Section 5 describes the FAETHON a/v archive interfaces. Finally, conclusions are drawn in Section 6.

2 System Overview

During its conception phase, the overall objective of FAETHON has been decomposed into five primary objectives, namely: (a) extraction of high level semantic information out of existing low-level semantic data, (b) generation and update of user profiles that drive and influence the weighting of the record sets that is queried out of the audiovisual archives, (c) incorporation of the retrieval of high-level semantic information and personalization into generic DBMS schemes, (d) development of intelligent indexing, querying and retrieval mechanisms for effectively handling the huge volumes of data within audiovisual archives and (e) exploitation of the availability of a/v archives, to fine-tune, evaluate, validate and integrate the project's results.

From the end-user point of view, this novel system exploits the advances in handling a/v content and related metadata, as introduced by MPEG-4, and MPEG-7, to offer advanced access services characterized by the tri-fold *semantic phrasing of the request*, *unified handling* and *personalized response*. From the technical point of view, FAETHON plays the role of an intermediate access server residing between end users and multiple heterogeneous archives, in order to:(a) keep track of user preferences,(b) project these preferences to appropriate indices of the archived content and (c) adjust the responses to user's queries in a manner that fits to his/her priorities.

FAETHON consists of two major components: the *FAETHON Intelligent Metadata Creation Package* that automatically generates content-related metadata to be used for browsing / search / retrieval, and the *FAETHON A/V Search Engine & Portal* that provides end user unified and personalized access to a number of heterogeneous a/v archives. It also contains a number of individual *a/v archive interfaces*, which are responsible for the MPEG-7 compliant communication between the FAETHON search engine and the participating archives, in a way transparent to the end user.

3 FAETHON Intelligent Metadata Creation Package

The FAETHON Intelligent Metadata Creation Package manages the FAETHON knowledge and user / archive profiles, analyzing the audiovisual material of a number of heterogeneous audiovisual archives and automatically generating material-related metadata in the form of a semantic index to be used by the FAETHON A/V Search Engine & Portal for unified access to the material. It consists of the following units:

3.1 Encyclopedia Editor

The FAETHON fuzzy relational encyclopedia consists of crisp and fuzzy sets, fuzzy relations and rules concerning the expansion of the fuzzy sets and relations [6,7]. An XML schema has been defined for this purpose, however, the general difficulties that pertain the manipulation of validated XML documents, as well as the need for an automation of the rules and the operation highlighted the necessity of a GUI-based tool.

This tool, the Encyclopedia Editor, has the following functionalities: (a) Graphical manipulation of fuzzy and crisp sets of semantic entities and fuzzy relations. Entities, crisp and fuzzy sets, relations and relation elements can be added, deleted and re-named. (b) Operations on fuzzy sets and fuzzy relations. (c) Fuzzy relation properties are visually listed for each relation. (d) Automatic expansion based on rules stored within the encyclopedia. (e) XML serialization. The encyclopedia editor uses XML to read and store the encyclopedia.

3.2 Description Graph Editor

The Description Graph Editor (DGE) is a graphic user interface that allows the definition of description graphs (DG) for a given semantic entity present in the FAETHON encyclopedia [8]. This definition can rely on previous simpler semantic entities or on MPEG-7 compliant visual descriptors. The system allows defining semantic relations between the semantic entities that form the DG.

The Description Graph Editor has the following additional functionalities: (a) Very simple interaction. (b) Definition of new semantic relations to be used in the creation of future DGs. (c) Definition of a Semantic Entity only based on MPEG7 visual descriptors. (d) Assignment of multiple Description Graphs to the same Semantic Entity, with different level of complexity. (e) Assignment of relevance values to the various nodes (simpler SEs and SRs) defining the Description Graph.

3.3 Detection of Events and Composite Objects in Video Sequences

The Detection of Events and Composite Objects (DECO) unit detects specific composite objects or events in video sequences and related metadata in order to extract their semantic interpretation [1]. It uses the semantic entity definitions (objects, events, concepts) stored in the FAETHON encyclopedia and scans the a/v documents

available at the participating archives in order to find and store (as links) the a/v documents that contain each semantic entity, together with a corresponding degree of confidence. Its output is stored in the FAETHON semantic index to be used by the search engine.

Events and composite objects are Semantic Entities (SEs), which are defined in the FAETHON Encyclopedia by means of the so-called Description Graphs (DGs) and interrelated using Semantic Relations (SRs). Detection and recognition of such entities in video sequences and related metadata is based on DG modeling and representation in order to capture their structure. A generic hybrid neuro-fuzzy architecture has been implemented, which provides rule-based inference and adaptation using the ideas of neural learning and fuzzy inference. The rules for SE recognition are closely coupled to their DG definitions, while state-of-the-art video processing/analysis techniques have been implemented for the detection of fuzzy predicates [9,10,11,12], i.e. extraction of the main mobile objects (and their interrelations) in video sequences, in the general case of sequences acquired by a moving camera. Built-in knowledge about entities and adaptation permits robustness and uncertainty handling, while learning allows updating of knowledge, i.e. adaptation to environmental changes.

3.4 Dynamic Thematic Categorization of Multimedia Documents

The Dynamic Thematic Categorization (DTC) unit performs automatic categorization of multimedia documents through matching between the textual metadata associated to the archived material and the thematic categories stored in the FAETHON encyclopedia. It unifies the thematic categories of the a/v archives and scans the a/v documents available at the participating archives in order to find and store (as links) the a/v documents that belong to each thematic category, along with corresponding degrees of confidence [7]. The output of the DTC is stored in the FAETHON semantic index and used for fast retrieval of a/v documents by the search engine.

The FAETHON categorization scheme classifies a document to one or more categories, according to their content [13]. Thus, documents belonging to the same category can be treated similarly, with respect to, e.g., user profiling, document presentation etc. DTC handles thematic categorization, i.e. a categorization in a conceptual level. This categorization uses the semantic entities encountered in a document, in order to classify the latter into classes, such as sports, diplomacy, chemistry and so on. The DTC unit performs a fuzzy hierarchical clustering of the semantic entities [14], relying on knowledge that is stored in the form of semantic relations. The notion of context has a central role in this process. Moreover, the DTC is able to perform categorization based on low-level audiovisual signal processing, using an approach based on decision trees.

3.5 User Profile Generation

Personalization of retrieval is the approach that uses the user profiles, additionally to the query, in order to estimate the users' wishes and select the set of relevant documents. In this process, the query describes the user's current search, which is the local

interest, while the user profile describes the user's preferences over a long period of time; we refer to the latter as global interest [15]. The User Profile Generation unit generates and updates user preferences based on the usage history. The usage history is updated after the end of a user query by storing all transactions of the user during the query process. The above transactions characterize the user and express his personal view of the a/v content. The user profile generation unit takes these transactions as input and with the aid of the encyclopedia 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 consist of a metadata preferences and a semantic preferences part. The metadata preferences refer to information like creation, media, classification, usage, access, and navigation preferences (e.g. favorite actors / directors or preference for short summaries). The semantic preferences, on the other hand, consist of preferences for semantic entities (interests), as well as preferences for thematic categories [16]. The generation of metadata preferences is implemented using a fuzzy hierarchical clustering and topic extraction technique; the metadata preferences are then used by the Presentation Filtering Unit. Both support unsupervised, fully automatic operation.

4 FAETHON A/V Search Engine & Portal

The FAETHON A/V Search Engine & Portal provides to the end user unified and personalized access to a number of heterogeneous audiovisual archives and supporting services including: semantic / metadata search to a/v material, browsing the material based on unified thematic categories, and personalized presentation of a/v material. It consists of the following units:

4.1 User Query Analysis Unit

The User Query Analysis Unit performs semantic analysis and interpretation of queries given by the end-user in the form of keyword expressions. Thus, it supports semantic phrasing of the user request in a high, conceptual level. It produces a semantic expression corresponding to each given keyword expression; this semantic expression consists of semantic entities (SEs), and degrees of confidence for each detected SE. In doing so, it uses knowledge stored in the FAETHON encyclopedia. It supports three operations: query interpretation, expansion and personalization.

During query interpretation, the keyword expression is transformed into a corresponding semantic expression, with keywords having been replaced by semantic entities and relations, and a corresponding degree of relevance. During query expansion, the context of the query is automatically detected and then the FAETHON thesaurus is utilized to map each semantic entity found in the semantic expression to a set of entities related to it in the specific context and expand the semantic expression with all the entities related to the initial entities [17,18]. Finally, during query personalization, user preferences are utilized so that the search process is "directed" towards fields in which the documents that satisfy the user request are most likely to be found.

4.2 A/V Document Search Engine

The A/V Document Search Engine is responsible for the identification of a/v documents matching the user query by both searching the FEATHON semantic index (semantic search) and communicating with the participating a/v archives (metadata search). For the semantic search, it utilizes the semantic expression produced by the user query analysis unit and produces a list of matching documents, along with the corresponding degrees of confidence [13]. For the metadata search, it constructs and dispatches a unified query to each participating archive, requesting the a/v documents that satisfy the metadata expression contained in the user query. The a/v archive interfaces are responsible for translating and handling this query. The two results produced by the semantic and metadata search from all archives are then combined and the result is a list of a/v document locators matching the user query. This process supports both semantic phrasing of each user request, and unified handling of the request in all archives, in a way transparent to the end user.

4.3 A/V Document Classification and Ranking Unit

The A/V Document Classification and Ranking Unit performs ranking (but not filtering) to the a/v documents retrieved by the search engine based on semantic preferences contained within the user profiles. Dynamic categorization 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 preferences [15]. The result is a list of a/v documents with their descriptions, ranked according to the interests that are stored in the user profile. Ranks are updated based on similarity measures, taking into account the degree of confidence of each user preference.

4.4 A/V Document Presentation Filtering Unit

The main purpose of the A/V Presentation Filtering Unit is to rank MPEG-7 formatted documents from a/v archives with regard to user preferences that are also stored in MPEG-7 format [7]. Additional functionality allows creating FAETHON specific responses that can be used by the FAETHON System. Through a well-defined and documented interface this component can be used for ranking of documents in other MPEG-7 related systems as well.

4.5 End User Interface

The End User Interface consists of the *User Interaction*, *User Presentation* and *User Communication* modules. The FAETHON *User Interaction* module offers a fully operational and functional interface to each user for completing a set of actions through a common web browser. The *User Communication Module* translates queries from the user interaction module according to the definition of the *user query* in the information model and parses the Faethon response for presentation purposes. The user can post a query using the form that consists of the semantic query and metadata

fields, such as title, date etc. The *User Presentation Module* dynamically generates search results from the user queries that are presented in several modes according to the personalization preferences as well as the user access rights. Then, the user has the possibility to see details of the retrieval process, i.e. the semantic entities that matched his textual query (Query Interpretation), the expanded list of semantic entities (Query Expansion), the re-ranked list after the impact of the user's profile settings and the final ranked list of the documents, as shown in Fig. 1.

SEMANTIC AND METADATA SEARCH - SemanticResponse

The expanded set of semantic entities has been matched with the following multimedia documents in the Faethon semantic index, with degree of relevance:

Id	Title	SourceArchive	Score
35	Flugzeugkatastrophe	FAA	0.9
1199	Επικαιρότητες Αυγούστου 1974	ERT	0.8
52	Die Vietnamkrise	FAA	0.78
1	Sensationell neuen Rettungsmethode	FAA	0.72
1514	Περισκόπιο	ERT	0.68
AVQ-A-004129-0038	Archeological excavations in Rome	Alinari	0.6
11	Ausbau und Elektrifizierung der Strecke Graz-Bruck	FAA	0.5
FCC-F-021960-0000	Exodus of the Belgian population	Alinari	0.45

Pages: << Previous | QueryInterpretation | QueryExpansion | SemanticResponse | PresentationResponse | ClassificationResponse | Next >>

Fig. 1. Ranked multimedia documents retrieved for a user query with the keyword "politics". The bar at the bottom indicates the intermediate steps.

5 FAETHON A/V Archive Interfaces

The individual a/v archive interfaces are responsible for the MPEG-7 compliant communication between FAETHON and the participating archive systems, in a way transparent to the end user. Three archive interfaces have been implemented for the three archives participating in FAETHON: FAA (Film Archive Austria), ERT (Hellenic Broadcasting Corporation) and Alinari Archive.

The *FAA archive interface* provides functionality for querying the publicly available data of FAA by SOAP via HTTP. The queries accepted by that interface are MPEG-7 formatted and are translated into the native format of the FAA system. The interface also includes a database access layer, which performs the connection to FAA's ORACLE database.

The ERT (Hellenic Broadcasting Corporation) archive utilizes the MPEG-7 content description standard and consists of two servers, namely (i) the database server, which hosts the archive database with all a/v content metadata, and (ii) the media server / web server / web service provider, which hosts all the a/v content itself, handles media streaming, provides an end-user web interface to the archive content, and serves as a web service provider to interface with the central FAETHON system. For the communication between ERT archive and FAETHON central system structured XML is used, which the *ERT archive interface* produces by parsing and translating the user query. Then an additional database access layer performs the search in ERT database, produces the response, assembles it into an MPEG-7 compliant format and returns it to FAETHON central system. This process makes access to the ERT archive transparent to the end user.

Finally, the *Alinari archive interface* establishes the communication between the FAETHON system and the Alinari Archive system. Additionally, Alinari interface poses the user query to the database and returns the results. Same as above, all communication between system's modules is achieved with MPEG-7 compliant format in XML, while the whole procedure of the communication of FAETHON with Alinari archive remains transparent to the user.

6   Conclusions

The key aspect of the FAETHON developments has been the generation and use of metadata in order to provide advanced content management and retrieval services. The Web will change drastically in the following years and become more and more multimedia enabled, making already complex content management tasks even more complex and requiring solutions based on Semantic Web technologies. Unlike today, content itself will be a commodity in a future Web, making the use of metadata essential. Content providers for instance will have to understand the benefits obtained from the systematic generation of metadata; service providers will have to accept metadata as the basis on which to build new services; and the producers of software tools for end-users will redirect their imagination towards more appropriate integration of application software with Web content, taking advantage of metadata. These developments clearly present some challenging prospects, in technological, economic, standardisation and business terms.

Another interesting perspective of FAETHON's developments is the personalisation, based on usage history, of the results of content retrieval. Personalisation software is still in its infancy, which means there are no turnkey solutions. Solutions using agent technologies still have a lot of hurdles to overcome. To improve this scenario, additional technology approaches need to be evaluated and areas of improvement identified. In both perspectives, clearly FAETHON made some interesting steps on the correct route and its developments are currently influencing the next research activities in the area of semantic based knowledge systems [19].

References

1. G. Stamou, Y. Avrithis, S. Kollias, F. Marques and P. Salembier, "Semantic Unification of Heterogenous Multimedia Archives", Proc. of 4th European Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2003), London, UK, April 9-11, 2003.
2. J. Hunter, "Adding Multimedia to the Semantic Web: Building an MPEG-7 Ontology", in Proc. of First Semantic Web Working Symposium, SWWS'01, Stanford University, California, USA, July 2001.
3. T. Sikora, "The MPEG-7 Visual standard for content description - an overview", IEEE Trans. on Circuits and Systems for Video Technology, special issue on MPEG-7, 11(6):696–702, June 2001.
4. ISO/IEC JTC1/SC29/WG11 N5231, "MPEG-21 Overview", Shanghai, October 2002.

5. Akrivas, G, Stamou, G, "Fuzzy semantic association of audiovisual document descriptions", Proc. of Int. Workshop on Very Low Bitrate Video Coding (VLBV01), Athens, Greece, October 2001.

6. Akrivas, G, Stamou, Stefanos Kollias , "Semantic Association of Multimedia Document Descriptions through Fuzzy Relational Algebra and Fuzzy Reasoning", IEEE Transactions on Systems, Man, and Cybernetics, part A, Volume 34 (2), March 2004.

7. Y. Avrithis, G. Stamou, M. Wallace, F. Marques, P. Salembier, X. Giro, W. Haas, H. Vallant, and M. Zufferey, "Unified Access to Heterogeneous Audiovisual Archives", Proc. of 3rd International Conference on Knowledge Management (IKNOW '03), Graz, Austria, July 2-4, 2003.

8. X. Giró, F. Marqués, "Semantic entity detection using description graphs", Workshop on Image Analysis for Multimedia Services WIAMIS-03, pp. 39-42, London, April 2003.

9. G. Tsechpenakis, Y. Xirouhakis and A. Delopoulos, "A Multiresolution Approach for Main Mobile Object Localization in Video Sequences", International Workshop on Very Low Bitrate Video Coding (VLBV01), Athens, Greece, October 2001.

10. Tsechpenakis, N. Tsapatsoulis and S. Kollias, "Probabilistic Boundary-Based Contour Tracking with Snakes in Natural Cluttered Video Sequences", International Journal of Image and Graphics: Special Issue on Deformable Models for Image Analysis and Pattern Recognition, to appear.

11. J. Ruiz and P. Salembier, Robust segmentation and representation of foreground key-regions in video sequences, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'01), Vol. III: IMDSP-L5, Salt Lake City, Utah, USA, May 7-11, 2001.

12. V. Vilaplana, F. Marqués, "Join detection and segmentation of human faces in color images", Proceedings of the EUROIMAGE ICAV3D 2001, International Conference on Augmented, Virtual Environments and three dimensional imaging, pp.347-350, Mikonos, Grecia, Mayo del 2001.

13. Wallace, M., Akrivas, G., Mylonas, P., Avrithis, Y., Kollias, S. "Using context and fuzzy relations to interpret multimedia content", Proceedings of the Third International Workshop on Content-Based Multimedia Indexing (CBMI), IRISA, Rennes, France, September 2003.

14. Wallace, M., Akrivas, G. and Stamou, G., "Automatic Thematic Categorization of Documents Using a Fuzzy Taxonomy and Fuzzy Hierarchical Clustering", Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), St. Louis, MO, USA, May 2003.

15. Wallace, M., Akrivas, G., Stamou, G. and Kollias, S., "Representation of user preferences and adaptation to context in multimedia content-based retrieval", Proceedings of the Workshop on Multimedia Semantics, SOFSEM 2002: Theory and Practice of Informatics, Milovy, Czech Republic, November 2002.

16. Wallace, M. and Stamou, G., "Towards a Context Aware Mining of User Interests for Consumption of Multimedia Documents", Proceedings of the IEEE International Conference on Multimedia (ICME), Lausanne, Switzerland, August 2002.

17. Akrivas, G., Wallace, M., Andreou, G., Stamou, G. and Kollias, S., "Context-Sensitive Semantic Query Expansion", Proceedings of the IEEE International Conference on Artificial Intelligence Systems (ICAIS), Divnomorskoe, Russia, September 2002.

18. Akrivas, G., Wallace, M., Stamou, G. and Kollias, S., "Context-Sensitive Query Expansion Based on Fuzzy Clustering of Index Terms", Proceedings of the Fifth International Conference on Flexible Query Answering Systems (FQAS), Copenhagen, Denmark, October 2002.

19. I. Kompatsiaris, Y. Avrithis, P. Hobson and M.G. Strinzis, "Integrating Knowledge, Semantics and Content for User-Centred Intelligent Media Services: the aceMedia Project", in Proc. of Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS '04), Lisboa, Portugal, April 21-23, 2004.