

A MEDIATOR SYSTEM FOR HETERO-LINGUAL AUDIOVISUAL CONTENT

Manolis Wallace

Department of Computer Science
University of Indianapolis, Athens Campus, Greece
wallace@uindy.gr

Thanos Athanasiadis, Yannis Avrithis, Giorgos Stamou and Stefanos Kollias

Image, Video and Multimedia Laboratory
Department of Electrical and Computer Engineering
National Technical University of Athens – Greece
{thanos,iavr}@image.ntua.gr {gstam,Stefanos}@cs.ntua.gr

ABSTRACT

In this paper, an integrated information system is presented that offers enhanced search and retrieval capabilities to users of hetero-lingual digital audiovisual (a/v) archives. This innovative 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 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 the queries that users may issue, and the context determined by user profiles.

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.

As a result, digital archiving of multimedia content including video, audio, still images and various types of documents has been recognized by content holding organizations in numerous countries as a mature choice for the preservation, preview and partial distribution of their assets. The advance 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 the last century, the attempt has been to develop intelligent and efficient human computer interaction systems, enabling the user to access vast amounts of heterogeneous information, stored in different sites and archives; these archives may be located in different countries. In order to achieve this objective metadata are attached to the original data, typically in the native language of the country where the archive is located.

Current and evolving international standardization activities, such as EBU, MPEG-4 [6][7][3], or JPEG2000 [5] for still images, deal with aspects related to data structures and metadata. In particular, the new MPEG-4 and MPEG-7 standards are object-oriented, i.e., adopt video objects as information units instead of scenes and shots. The MPEG-7 standard, formally named "Multimedia Content Description Interface", provides a rich set of standardized audiovisual Description Tools (the metadata elements and their structure and relationships, that are defined by the standard in the form of Descriptors and Description Schemes) to describe multimedia content.

Content Description DSs are classified into structural and conceptual (semantic) DSs. The structural DSs provide a low-level and machine oriented kind of description, while conceptual DSs express a high-level and human oriented kind of description. The use of the conceptual description for searching in multimedia databases has advantages over structural description, because of its proximity to human understanding of multimedia information. Moreover, a query based on semantic entities can be personalized, taking into account the individual user's interests and preferences, which cannot be directly included in a structural query.

The FAETHON system [2] performs a semantic unification of different archives by using an encyclopaedia which contains definitions of abstract classes. Creation of the encyclopaedia relies both on human experts and existing ontologies. The system correlates the specific semantic entities of the multimedia descriptions of the individual archives to the abstract ones of the encyclopaedia. When a user makes a query, the supplied keywords are translated into the semantic entities of the encyclopaedia. The documents whose descriptions have been correlated to the requested semantic entities are retrieved, and then filtered and ranked by taking into account the user's preferences in each semantic entity. By including concept definitions in different languages in the encyclopaedia, and then

choosing the set of definitions to use for each archive based on its location, we achieve a unified semantic indexing of archives, regardless of the language used in their annotation with metadata. In this way, the FAETHON system is able to provide for unified access to hetero-lingual audiovisual archives.

ARCHITECTURE

From a technical point of view, the overall functionality of the proposed system that of a mediator (to offer a unified interface that will allow end-users to have efficient access to a number of individual audiovisual archives).

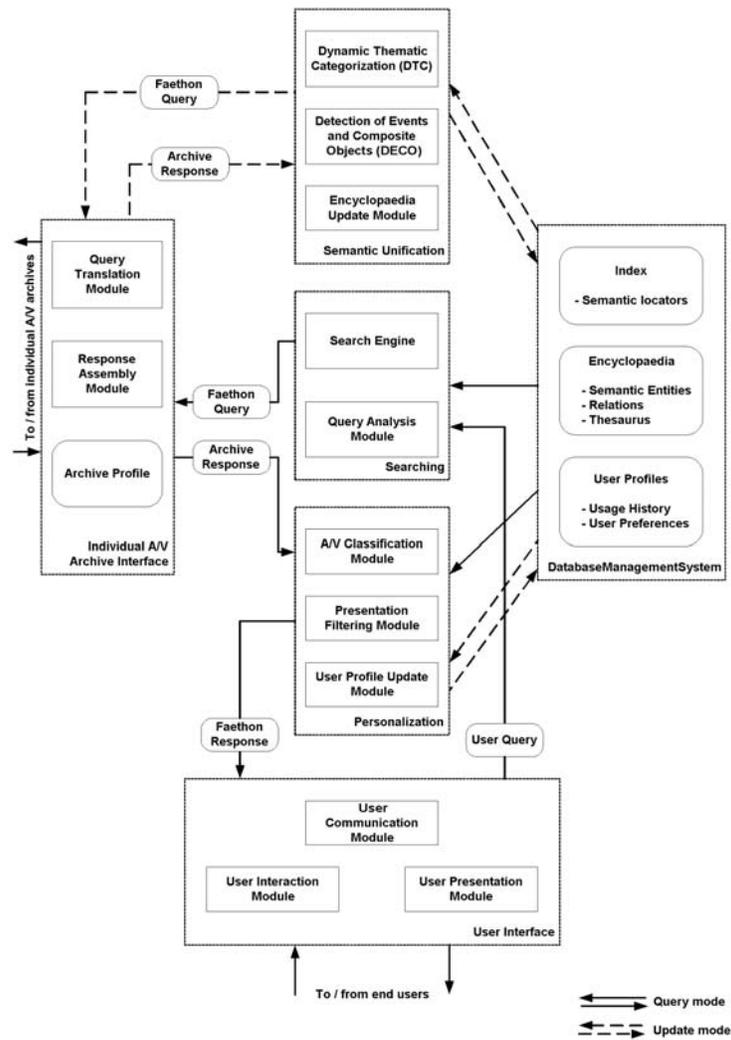


Figure 1. The FAETHON system

In this sense FAETHON, acts as an intermediate agent that undertakes the tasks of (1) receiving end-users queries, (2) translating the terms of the query in a set of semantic

entities by means of the knowledge of the system, (3) searching the audiovisual archives for the existence of the above semantic entities, (4) receiving the produced responses and (5) presenting the latter to the end-user in order of importance, ranking them using the user profile. The above procedure is the typical information flow in a 3-tier environment.

What is innovative in this flow is the semantic level of the resolution of the users' queries. Based on the above process, FAETHON's users are able to issue expressive semantic queries whose answers comprise "understanding" of the involved semantic entities and rapidly converge to the focus, i.e. to what the end user has in mind, understanding the context of the query by also using the information of the user profile [4].

The FAETHON system enables a user to perform a single query on multiple hetero-lingual multimedia archives and receive the results in a uniform manner. Its operation has two distinct modes: the query and the update mode (working in parallel). In the query mode FAETHON system serves its end-users by exploiting (a) its already available knowledge, (b) pre-processed information previously extracted from the audiovisual archives and (c) the on-line access to the latter using the user profile and relevance feedback. The continuous arrows in Figure 1 present the system operation in query mode. In its update mode, FAETHON system enhances its knowledge and gathers information from the audiovisual archives, processes this acquired information and stores it for subsequent use. Moreover, it updates user profiles translating the usage history into user preferences. The dotted arrows in Figure 1 present the system operation in update mode.

THE KNOWLEDGE OF THE SYSTEM

The knowledge of FAETHON contains the encyclopaedia and the user profiles. Among other actions, it allows:

1. Structured storage of semantic entity descriptions and relations defined by experts.
2. Forming complex concepts and events (composite entities) by the combination of simple ones through a set of relations.
3. Expanding the user query by looking for related concepts to those words contained in the semantic part of the query.

Three types of information are included in the encyclopaedia, providing the information needed for these actions: semantic entities (SEs), semantic relations (SRs) and the thesaurus.

Semantic entities are entities such as objects, events, concepts, thematic categories, agents and semantic places and times. Each semantic entity can contain textual annotation, including keywords in various languages defining the entity, sub-entities and their relations and low-level descriptors. A special kind of semantic entity is the thematic category, which corresponds to concepts such as 'sports' or 'news reports'. The purpose of this special semantic entity is to provide the context for a user's query.

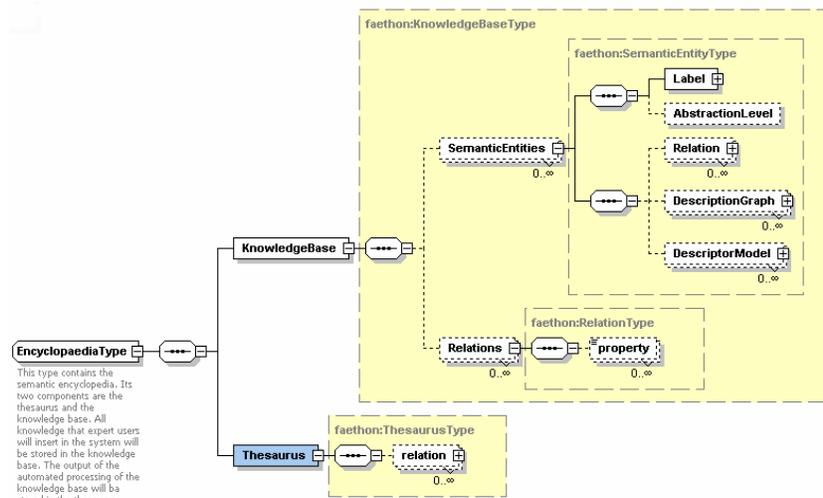


Figure 2. Graphical representation of the encyclopaedia type

Semantic relations are the relations linking related concepts as well as the relations between simple entities to allow forming composite ones. The description of all relationships among the SEs in the encyclopaedia, using the semantic relations, forms a graph structure. The graph nodes correspond to the SEs, whereas graph links represent the type of relationship between the nodes connected by them. This graph structure is represented in FAETHON by means of a SemanticEntities DS and a SemanticRelations DS. All relations are, in principle, fuzzy, and each relation value can be any number between 0 and 1. In practice, all relations are sparse (each entity is related to only a small number of entities), and are represented by a sparse matrix [13].

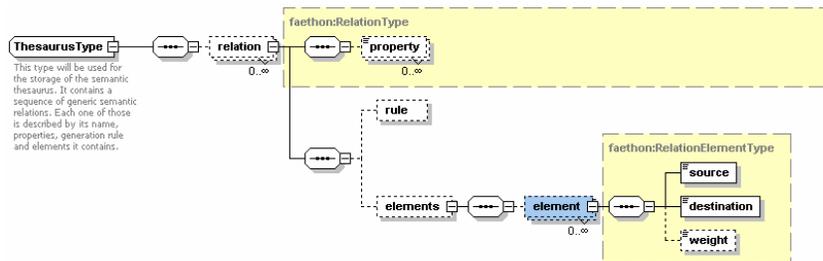


Figure 3. Graphical representation of ThesaurusType

The thesaurus provides simplified views of the knowledge that is contained in the encyclopaedia, to be used for specific tasks. Thus, the thesaurus contains relations that are intended to facilitate query expansion, personalization, detection of thematic categories etc. The thesaurus is generated automatically from the semantic relations of the encyclopaedia. The concept of Thesaurus is unique to FAETHON and, therefore, DSs additional to those of MPEG-7 are specified.

SEMANTIC UNIFICATION AND SEARCH

The role of semantic unification is to correlate the multimedia document descriptions provided by the archives with the semantic entities stored in the encyclopaedia. The result is, on one hand, the semantic index, containing the correlations between multimedia documents and semantic entities, and on the other hand, the thesaurus. The semantic unification is performed with the aid of the Detection of Thematic Categories (DTC) and the Detection of Events and Composite Objects (DECO) modules (Figure 1).

The DTC module, which is further presented in section “Thematic Categorization of Documents”, maps the description of a multimedia document to the set of thematic categories, and stores the relevance values in the semantic index. The DECO module maps the description of a multimedia document to the set of semantic entities. It is similar in operation to the DTC module, but it is capable of matching composite semantic entities (entities that contain sub-entities) to sets of consecutive DSs found in the description of the document [15].

The semantic index produced by DTC and DECO provides a fuzzy mapping between the set of the semantic entities of the encyclopaedia and the set of the document locators of the documents of all the archives. It therefore contains degrees of relevance between semantic entities and documents. It is used to locate documents that match the user query without searching in the archives at query time. The searching procedure takes as input the keywords that consist of the semantic part of the user's query and the metadata the user has provided. The processing of the user's query consists of the query interpretation and the query expansion phases. In query interpretation, each keyword is transformed into a fuzzy set defined on the set of semantic entities. On the other hand, in query expansion, the above sets are expanded using the information of the fuzzy thesaurus. Finally, the search engine uses the semantic entities involved in the expanded query and returns the associated document locators based on the information of the semantic index.

Thematic Categorization of Documents

The intelligent module of DTC accepts as input the Semantic Index. This is in fact a fuzzy relation between documents and semantic entities. Moreover, the semantic index must be a normal fuzzy set for each document. Based on this relation, and the knowledge contained in the available semantic relations, the module aims to detect the degree to which a given document is related to a thematic category. In other words, the module attempts to calculate a fuzzy relation between documents and thematic categories [9][14].

In designing an algorithm that is able to calculate this relation, in a meaningful manner, a series of issues need to be tackled:

1. A semantic entity may be related to multiple, unrelated thematic categories.
2. A document may be related to multiple, unrelated thematic categories.

3. The semantic index may have been created in an automated manner. Thus, existence of random, and therefore misleading semantic entities cannot be excluded.
4. Semantic relations are always a matter of degree. Therefore, correlation between a document and a thematic category is also a matter of degree.

According to issue 1, it is necessary for the algorithm to be able to determine which thematic categories are indeed related to a given document. In order for this task to be performed in a meaningful manner, the common meaning of the remaining entities that index the given document needs to be considered as well. This is accomplished through the consideration of the context of the document [9]. On the other hand, when a document is related to more than one, unrelated thematic categories, as issue 2 points out, we should not expect all the terms that index it to be related to one another, or to each one of the thematic categories in question. Quite the contrary, we should expect most entities to be related to just one of these thematic categories. Therefore, a clustering of semantic entities, based on their common meaning, needs to be applied. In this process, entities that are misleading will probably not be found similar with other entities that index a document. Therefore, the cardinality of the clusters may be used to tackle issue 3. Finally, issue 4 is easily solved by allowing the algorithm to be fuzzy.

PERSONALIZATION

The FAETHON system is designed to simultaneously access several hetero-lingual audiovisual archives. Thus, when expressing a query, users may get thousands of matches as a response; this is essentially a result of the increase in detailed metadata information that will accompany the a/v content. So a personalized view on the query result is an important issue for the system.

User related information is fundamental to the personalization of user queries and archive responses [8]. FAETHON supports both passive and active user profiling. The former is an automatic update of the user profile, which is further analyzed in section “Extraction of User Preferences”, whereas the latter requires active user involvement. The user profile consists of two major parts: the user preferences and the user history, which contains the information relative to the user-FAETHON system interactions

During the registration process, the user manually specifies his profile settings consisting of metadata and semantic related preferences. For each preference, he defines a weight indicating his interests (like/dislike value).

During the presentation process a ranking of the retrieved records is performed by using the user semantic preferences (audio/visual classification module) and the user metadata preferences (presentation filtering module). Each module produces independently a ranked list of documents, which are successively merged by taking into account the weighting performed by these modules and the importance of the each module itself within the personalization process.

Extraction of User Preferences

Based on the analysis of DTC, FAETHON detects the topics that are related to each one of the documents in a user's history. Still, this does not render the problem of semantic user preference extraction trivial. What remains is the determination of the following:

1. How should a user be modeled?
2. Which of these topics are indeed of interest to the user, and which are found in the history due to coincidental reasons?
3. To which degree is each one of these topics of interest to the user?

As far as the user model is concerned, main principles may be summarized in the following [10]: (1) special care must be taken for the representation of negative preferences, (2) it is necessary to store negative preferences separately from positive ones, so that they may be processed separately, (3) each positive interest needs to be stored separately. As far as detection of preferences is concerned, the main points to consider may be summarized in the following: (1) A user may be interested in multiple, unrelated topics, (2) not all topics that are related to a document in the usage history are necessarily of interest to the user.

These are tackled using similar tools and principles, as the ones used to tackle the corresponding problems in detection of thematic categories. Thus, once more, the basis on which the extraction of preferences is built is the context. The common topics of documents are used in order to determine which of them are of interest to the user and which exist in the usage history coincidentally. Extraction of metadata preferences is based on similar principles [11].

FAETHON A/V SEARCH ENGINE & PORTAL

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.

The FAETHON A/V Search Engine & Portal provides to the end user seamless, unified and personalized access to all of the abovementioned archives, although different languages have been used in annotating their documents. 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.

| 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 | Επικραδότητες Αυγούστου 1974 | ERT | 0.8 |
| 52 | Die Vietnamkrise | FAA | 0.78 |
| 1 | Sensationell neuen Rettungsmethode | FAA | 0.72 |
| 1514 | Περικόπτιο | 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 >>

Figure 4. Ranked multimedia documents retrieved for a user query with the keyword “politics”. The bar at the bottom indicates the intermediate steps.

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 [16][17]. 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.

In Figure 4 we present a sample output of the system's user interface. One can see that results from all archives appear in a uniform form that seamlessly hides from the user

the fact that the annotation languages are fundamentally different (Greek for ERT, German for FAA and English for Alinari).

CONCLUSIONS

This paper presented an integrated information system that offers enhanced search and retrieval capabilities to users of hetero-lingual digital audiovisual archives. The proposed system focuses on intelligent extraction of semantic information from audiovisual and text related data taking into account the nature of the queries that users may issue, and the context determined by user profiles. It also provides a personalization process of the response.

All of the system's intelligent operations are based on a novel encyclopaedia that contains both the definitions of abstract concepts and specific events and objects (semantic entities), as well as the relations that exist among them. A first version of the FAETHON prototype, with a limited encyclopaedia, has been developed and is currently undergoing testing.

Future work includes extension of the encyclopaedia, as to cover more topics and to a greater extent. Moreover, more work shall be done towards the integration of the system with a robust implementation of the DECO module. More information on the status and goals of FAETHON can be found at the project's home page (<http://www.image.ece.ntua.gr/faethon/>).

ACKNOWLEDGMENTS

This work was partially funded by the EC IST-1999-20502 Project. This work was partially funded by the HERACLETUS fund for basic research.

REFERENCES

1. Akrivas G. and Stamou G.: "Fuzzy Semantic Association of Audiovisual Document Descriptions", in Proc. of Int. Workshop on Very Low Bitrate Video Coding (VLBV), Athens, Greece, October 2001
2. Avrithis Y. and Stamou G.: "FAETHON: Unified Intelligent Access to Heterogeneous Audiovisual Content", in Proc. of Int. Workshop on Very Low Bitrate Video Coding (VLBV), Athens, Greece, October 2001
3. "Text of 15938-5 FDIS Information Technology - Multimedia Content Description Interface - Part 5 Multimedia Description Schemes", ISO/IEC JTC 1/SC 29/WG 11/N4242, 2001-10-23
4. Ganter B. and Wille R.: "Formal Concept Analysis: Mathematical Foundations", Springer-Verlag, 1999
5. "JPEG 2000 Part I Final Committee Draft Version 1.0", ISO/IEC JTC1/SC29/WG1 N1646R, 2000
6. Koenen R.: "Overview of the MPEG-4 Standard", ISO/IEC JTC 1/SC 29/WG 11/N4668, March 2002

7. Manjunath B. S., Salembier P. and Sikora T.: "Introduction to MPEG-7", Ed. John Wiley & Sons, West Sussex, P019 1UD, England, 2002
8. Soltysiak S. J. and Crabtree I. B.: "Automatic learning of user profiles - towards the personalisation of agent services", BT Technol J Vol 16 No 3 July 1998
9. 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.
10. 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
11. 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 and Expo (ICME), Lausanne, Switzerland, August 2002.
12. Wallace M., Avrithis Y., Stamou G., Kollias S. "Knowledge-based Multimedia Content Indexing and Retrieval" Multimedia Content and Semantic Web: Methods, Standards and Tools, Stamou G., Kollias S. (Editors), Wiley, 2004.
13. Wallace M., Kollias S., "Computationally efficient incremental transitive closure of sparse fuzzy binary relations", Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Budabest, Hungary, July 2004.
14. 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
15. Wallace M., Avrithis Y., Stamou G., Kollias S., Marques F., "Combining Raw Media and Metadata Information for the Understanding of Multimedia Content", submitted to the special issue on Analysis and Understanding for Video Adaptation.
16. 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
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