TOWARDS UNIVERSAL AND PERSONALIZED ACCESS TO AUDIOVISUAL CONTENT IN THE DYMAS SYSTEM*

Luis Herranz and José M. Martínez
Grupo de Tratamiento de Imágenes, Escuela Politécnica Superior
Universidad Autónoma de Madrid, E-28049 Madrid, Spain

ABSTRACT

This paper presents the current status of the Deferred Time Environment (DTE) of the DYMAS system. Currently, the DTE prototype enables seamless and personalized access to multimedia content over fixed networks (specifically over Internet). Integration of mobile phones access (Java and MMS terminals) is currently on-going. The system uses the framework of MPEG-7 and MPEG-21 to provide description metadata of the multimedia content and the usage context (including terminal, network capabilities and user preferences) respectively. These descriptions are the base for the main functionalities of the Annotation, Search and Browsing subsystems. A novel XSLT based architecture is proposed to proccess all the XML descriptions. The Streaming subsystem uses these metadata to configure a real-time transcoding session to deliver the multimedia content to the user over a commercial streaming service.

1. INTRODUCTION

Universal Multimedia Access (UMA) [1] refers to the capability of access to rich multimedia content through any client terminal and any network. The development of new wireless networks, providing multimedia capabilities, and a wide and growing range of client terminals makes the adaptation of content an important issue in future multimedia services.

One of the main objectives within the DYMAS project is the provision of alternative multimedia services based on content broadcasted in digital television channels. These services rely on the UMA concept and associated technologies and standards.

Different authors have published about general issues and architectures for UMA systems [1][2][4][5], but there are not so many papers about prototypes, test-beds, or implementations (e.g., [6][7]). In this paper we present the current status of a system enabling UMA using both MPEG-7 and MPEG-21 (in [6] the system uses only MPEG-21, whilst in [7] only MPEG-7 was used).

Section 2 introduces the DYMAS system; Section 3 presents the current architecture of the subsystem enabling UMA within DYMAS. Sections 4 and 5 describe, respectively, the different descriptions used and their relation with the MPEG-7 and MPEG-21 specifications, and the functionalities of the user applications. Section 6 concludes the paper.

2. OVERVIEW OF THE DYMAS SYSTEM

Figure 1 depicts an overview of the DYMAS System. It mainly describes a processing system with one information input (a DVB Transport Stream) and two information outputs (the modified DVB-TS and audiovisual services directed to other alternative access networks).

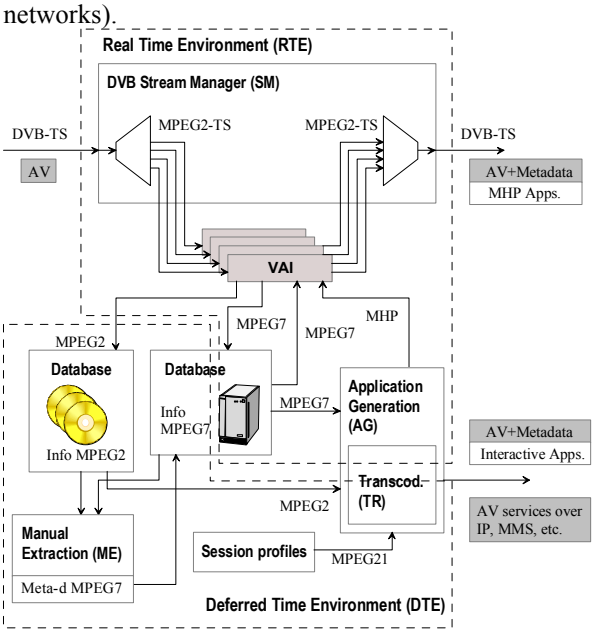

Figure 1. Block diagram of the DYMAS System

The system relies on technology for automatic content extraction from audiovisual information, which is currently immature, highly resource consuming, and just able to cope in real-time with low level basic features. These features are the basis for on-line service provision, that is, for the Real-Time Environment (RTE) [8].

However, in order to be open and scalable and to progressively adapt to research results, the framework here presented also considers the provision of services

* Work partially supported by the Ministerio de Ciencia y Tecnología of the Spanish Goverment under Project TIC2002-03692. Part of the work of Luis Herranz was partially supported by Telefónica de España under a grant of the "Cátedra Telefónica para Internet de Nueva Generación" of the Universidad Politécnica de Madrid.

that do not have a real-time requirement, but can conversely be offered with some delay. This is a responsibility of the Deferred-Time Environment (DTE). The DYMAS system uses the framework of MPEG-7[9], currently mainly the Multimedia Description Schemes [10] specification, to provide description metadata of the multimedia content. Some parts of the descriptions are generated in the RTE and used for providing added value services over DVB-MHP. Besides their use in interactive television applications, these descriptions are stored in the MPEG-7 database where descriptions are enriched via manual annotation and additional (non real time) automatic and supervised features extraction algorithms. These enhanced descriptions are the base for the UMA services provided by the DTE. Additionally the DTE makes use of MPEG-21[11], mainly the Digital Item Adaptation [12] specification, to provide description metadata of the usage context (including terminal, network capabilities and user preferences) in order to perform content adaptation.

3. CURRENT DTE ARCHITECTURE

The current DTE provides universal and personalized access to the MPEG-2 database over the Internet[13] (in the next months the access via mobile Java and MMS terminals will be also available). The DTE system architecture is shown in Figure 2. At the current status of the system, the applications run over Web Browsers running on standard PCs. There are two subsystems, the Metadata subsystem and the Streaming subsystem .

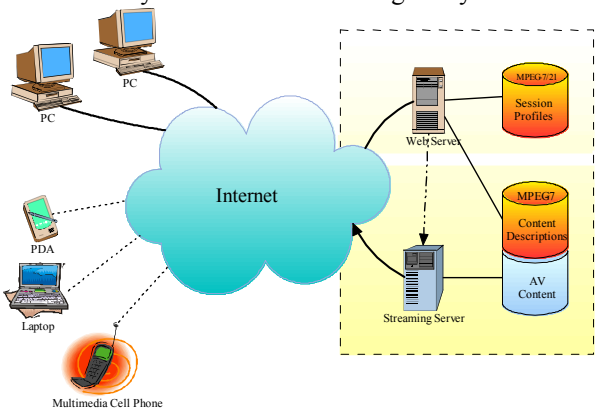

Figure 2. Current DTE Architecture

3.1. Metadata Subsystem

This part of the system (see Figure **3**) consists of a set of web applications: indexing/annotation, search and browsing.

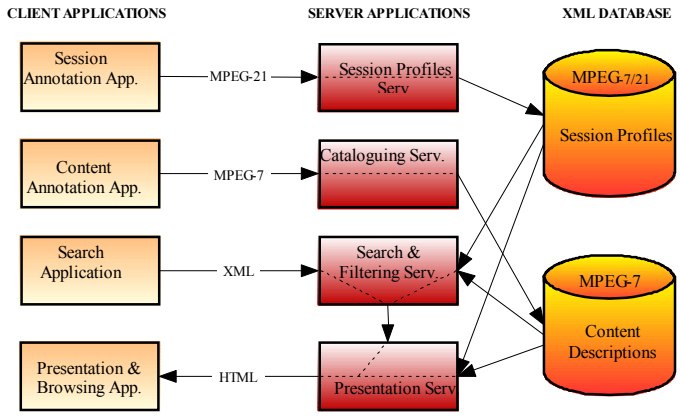

Figure 3. Metadata subsytem

- The Annotation application (see Figure 4) provides an easy interface to edit XML descriptions using a set of MPEG-7 and MPEG-21 description tools. These descriptions can refer to the multimedia content (according to MPEG-7 schema) and be used to catalogue new multimedia content, or can be descriptions of the session and usage context, including user preferences, terminal and network descriptions (according to MPEG-21 DIA schema, plus MPEG-21 Digital Item Declaration). At server side, new descriptions are stored on the corresponding description database when uploaded by users.

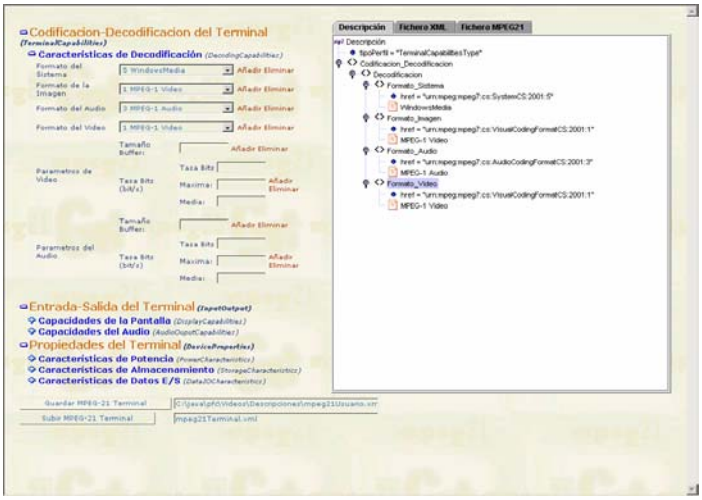

Figure 4. Annotation application screen

- The Search application provides an automatic search including (simultaneously or not depending on user selection) an automatic filtering based on the user preferences (part of the usage profile) and a query driven search

using the same GUI as the Annotation application. When the search results are available, the server sends a results HTML page to the client terminal.

- The Browsing application allows users to select content among the results page and access to a detailed description of the content with an embedded interface for playing the content.

3.1.1. XSLT-based architecture

In contrast with other XML processing models, the web system takes a new approach of processing based on XSL Transformations. In fact, it could be understood as a set of chained transformations, from client request (in XML/MPEG-7 format) to server response (in HTML), including search, filtering and presentation processes. XSLT pattern matching allows search engine process in a natural way the XML query, transforming it in an output XML document with the results of the search task. In the same transformation, the filtering is performed making use of the user profile description. At last, results are transformed one more time to format the results in order to send the response back to the user.

With this approach, the complexity and efficiency relies on the XLST processor, avoiding additional design of databases and complex optimized queries. XSLT processors are optimized to process XML data in parallel, and chained transformations can process nodes even when its previous transformation hasn't finished.

3.2. Streaming Subsystem

Using the user's session profile, the transcoding application of the streaming system configures a streaming session. The commercial streaming server (currently Windows Media Encoder 9) transcodes, based on some parameters calculated from the terminal and network capabilities parameters obtained from the session profile, the selected multimedia content. In the client side, the stream is played using a commercial player (currently Windows Media Player 9).

4. CONTENT, SESION AND QUERY DESCRIPTIONS

Within the system, we consider three descriptions: the content description, the session description, and the query description.

The content description uses MPEG-7, and is based on the MPEG-7 Simple Profile[14]. Content descriptions are created using the corresponding part of the Annotation application.

The session description uses MPEG-21 DIA description tools. The session description is split into three subdescriptions (MPEG-21 compliant), depending on the context element to be considered. A session description contains links to a network description, a terminal description and a user description (see Figure 5). Session descriptions are created using the corresponding part of the Annotation application (in a future, network and terminal will be detected automatically and user preferences updated automatically inspecting -if allowed- user's usage history). In order to make easier and faster the edition of session descriptions, the application allows users to select among previous terminal and network descriptions stored in the session descriptions database. Only when a terminal or network is not available in database, the user can edit a new description and submit it to server (in a commercial implementation, only authorized users –e.g., service provider- would be allowed to submit new terminal and network descriptions). This new description can be reused by future users. Even preferences can be reused (if allowed by privacy issues) to configure another session profile for the same user but using different terminals and network (until automatic terminal and network detection become available in the system).

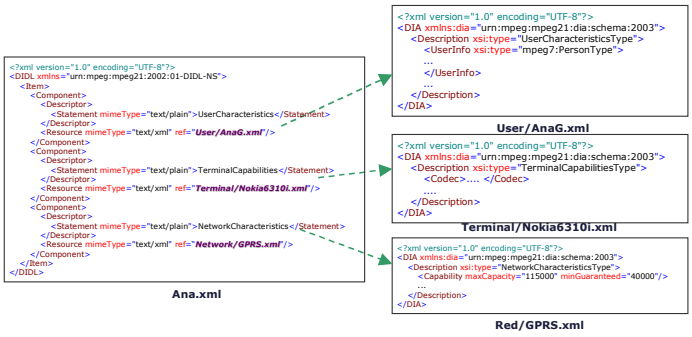

Figure 5. Session description and subdescriptions

The query description is an XML document, allowing the implementation of the search (filtering and query driven) capabilities using the XLST based architecture presented above. The query description uses a modified MPEG-7 profile which represents a partial content description that the user wants to match in the database. Therefore these "queries" are also XML descriptions similar to MPEG-7 content descriptions. However, as MPEG-7 is not designed for query, the query description is compliant with a modified MPEG-7 schema which includes some extra attributes (e.g., *case-sensitive, just-included*), and unconstraints some description elements.

5. ACCESING CONTENT

5.1 Search and automatic filtering

The search and filtering engine is based on a XSLT transformation, which matches paths in the XML query description with every MPEG-7 content description stored in the database (e.g.

../CreationInformation/Creation/Title values in the query description and the content descriptions).

Depending on the matching rule defined for each path (corresponding to a different descriptor), the search or filtering process can be adapted for each specific MPEG-7 description tool. One important new parameter defined for each matching rule is *relevance,* which sets the relative importance of a descriptor against other (e.g. a matching in a title is much more important than a matching of a word in a review annotation). The *relevance* value for each element is coded in the search and filtering XSLT transformation. Since XSLT files are plain text, the system administrator can edit the file and easily change relevance values, based on user feedback about his/her satisfaction with the results against the expected ones.

Users can store their search and filtering preferences in their session profiles (using MPEG-7 FilteringandSearchPreferences description tool) stating a preference (using MPEG-7 preferenceValue attribute) value for every query element for which a value is entered. In contrast with the *relevance* parameter, which is fixed by the system administrator for each different search descriptor, the preferences values are set by the user in his/her profile for each couple "field (descriptor) and value", and is only used if the automatic filtering is enabled.

For each query annotation (every query element for wich a value is entered) that match the corresponding descriptor value in any stored description, final relevance is computed as the sum of search relevance and filtering relevance (if filtering is enabled). Filtering relevance is computed from both relevance and preference values. Filtering information is used to refine the search classifying results according to user preferences, or even performing automatic information retrieval with no search query at all (if the user does not annotate any query field).

5.2 Personalized presentation of results

When search and filtering have finished, the search results are listed in a XML document. The last transformation in the XSLT chain transforms the list of results from neutral XML to personalized HTML. This transformation is described in a XSL file, defining a presentation template. Using different XSL-based presentation templates we can personalize the presentation to the client. Currently, these templates are created from HTML template pages, adding XSLT code manually. These templates are stored in the server and retrieved when user logs in the system. The name of the template to be used by a user is annotated in his user preferences description in the session description.

Figures 6 and 7 show two different personalized presentations, where we can see that depending on the personalization template, formatting, layouts and languages are different, and even the amount of the displayed information can be customized.

Transformations can be used to easily extend the web user interface (of the browsing application, which is generated by XSL transformations) to other terminals and applications. The XSLT processor can transform the XML results document into other markup languages, such WML, XHTML, etc., or even plain text, providing an easy way to generate an adapted page to other terminals, such as PDAs, mobile phones, etc. Future versions of the system will investigate further the implementation of this approach.

.

5.3 Automatic stream adaptation

The adaptation of the multimedia content is provided making use of MPEG-21 DIA description tools. Based on the network and terminal descriptions, the Streaming subsystem computes an optimized set of transcoding parameters according to the network resources and to terminal capabilities. These transcoding parameters (e.g. audio bitrate, sample rate, frame size, video bitrate) are used to configure a session in the streaming server, which transcodes the video file from content database into a new adapted stream. Users view this content embedded in the detailed information page (current presentation style).

5.4 Accesibility features

In order to extend the UMA capabilities to people with impaired hearing, we are incorporating a new feature in the system which allows to add closed captions to the media (see Figure 6 –bottom left-). At this moment, this capability is provided using an external XML captions file (Microsoft SAMI format) referenced from the content description (using the RelatedMaterial description tool).

Future versions will use MPEG-7 Spoken Content DS to provide synchronized text captions into the content description file after automatic speech recognition. The SAMI file will be created using another XSLT transformation from the SpokenContent DS description. We are also evaluating alternative formats (namely MPEG-4) for representing the results of the MPEG-7 description transmoding, including or referencing the content.

Besides accessibility issues, closed captions are useful in many specific fields, such as learning applications, karaoke systems, etc.

6. CONCLUSIONS

We have presented the current features and status of the DTE of the DYMAS system, in charge of providing

alternative access to multimedia content broadcasted in digital television channels. The system makes use of concepts, technologies and standards related to UMA, validating their functionalities and usability. The current system include the annotation applications allowing content cataloguing and a search and browsing application based on web browsers. Most of the developed architecture and software form the base of the additional access services which will be integrated shortly (mobile and wireless access via Java and MMS terminals).

The paper details the implementation of the main functionalities of the system based on MPEG-7 and MPEG-21. These functionalities include automatic search and filtering based on user preferences, personalized presentation and automatic content adaptation to user's context (terminal and networks).

7. REFERENCES

[1] Vetro, C. Christopoulos, T. Ebrahimi (guest editors), "Universal Multimedia Access (special issue)", *IEEE Signal Processing Magazine,* 20 (2), March 2003.

[2] R. Mohan, J.R. Smith, C.-S. Li, "Adpating Multimedia Internet Content for Universal Access", IEEE Transactions on Multimedia, 1(1):104-114, March 1999.

[3] J.R. Smith, "Universal Multimedia Access", in *Proc. SPIE Multimedia Systems and Applications IV*, vol. 4209, Nov. 2000.

[4] A. Perkis, Y. Abdeljaoued, Ch. Christopolous, T. Ebrahimi, J. Chicharo, "Universal Multimedia Access from Wired and Wireless Systems", *Birkhauser Boston Transactions on Circuits, Systems and Signal Processing (special issue on Multimedia Communications)*, 20(3):387-402, 2001.

[5] E. Fossbakk, P. Manzanares, J.L. Yago, A. Perkis, "An MPEG-21 framework for streaming media", in *Proc. of IEEE Multimedia Signal Processing 2001*, pp. 147-152, October 2001.

[6] A. Perkis, J. Zhang, T. Halvorsen,J.O. Kjode, F. Rivas, "A streaming media engine using digital media adaptation", in *Proc. of IEEE Multimedia Signal Processing 2002*, pp. 73-76, December 2002.

[7] J.M. Martínez, C. González, O. Fernández, C. García, J. de Ramón, "Towards Universal Access to Content using MPEG-7", *Proceedings of the ACM Multimedia 2002 Conference*, pp. 199-202.

[8] J. Bescós, J.M. Martínez, N. García, "Real-time audiovisual feature extraction for on-line service provision over DVB streams", *Visual Content Processing and Representation-VLBV03, LNCS Vol. 2849*, Springer Verlag, 2003, pp. 141-148.

[9] J.M. Martínez, R. Koenen, F. Pereira, "MPEG-7: the generic multimedia content description standard", *IEEE Multimedia*, 9(2): 78-87, April-June 2002

[10] MPEG MDS Group, "MPEG-7 Multimedia Content Description Interface, Part 5: Multimedia Description Schemes IS", ISO/IEC N4242, July 2001.

[11] I. Burnett, R.van de Walle, K. Hill, J. Bormans, F. Pereria, "MPEG 21: Goals and Achievements", IEEE Multimedia, 10(4):60-70, Oct.-Dec 2003.

[12] MPEG MDS Group, "MPEG-21 Multimedia Framework, Part 7: Digital Item Adaptation FCD", ISO/IEC N5845, July 2003.

[13] L. Herranz, *Acceso Multimedia Universal mediante MPEG-7 y MPEG-21 a través de terminales y redes fijas (Universal Multimedia Access using MPEG-7 and MPEG-21 through fixed terminals and networks)*, Master Thesis, E.T.S.I.T, Universidad Politécnica de Madrid, July 2003.

[14] MPEG Requirements Group, "MPEG-7 Profiles Under Consideration", ISO/IEC N6039, October 2003.

Figure 6. Browsing application screen

Figure 7. Search results screen

Figure 8. Search results screen with different presentation template