

USING METADATA TO PROVIDE SCALABLE BROADCAST AND INTERNET CONTENT AND SERVICES

GABRIELLA KAZAI^{1,2}, MOUNIA LALMAS¹, MARIE-LUCE BOURGUET¹ AND ALAN PEARMAIN²

*Department of Computer Science¹ and Department of Electronic Engineering²
Queen Mary University of London*

The SAVANT project is developing scalability and synchronisation techniques for delivering multimedia content over a variety of broadcast and IP networks to a variety of user devices. This paper deals with the use of metadata standards such as MPEG-7 and MPEG-21 to drive the automatic adaptation and personalisation of scalable broadcast and Internet content and services.

1. Introduction

The convergence of broadcasting and the Internet, the growth of digital TV and the increasing user mobility are the main driving forces behind the European IST SAVANT (Synchronised and Scalable AV content Across NeTworks) project. SAVANT [1] aims to simultaneously employ broadcast and telecom networks to achieve added-value services for digital and interactive TV, using *scalable content* and *scalable services* to allow end users to retrieve interrelated and synchronised multimedia content on terminals with different capabilities under varying network conditions. This scalable approach implies the automated adaptation of content and services to different user terminals and user preferences. It also supports the intelligent distribution of content based on the capabilities of the networks present. The adaptation of content and services relies on content and media management, and semantics-based annotation, delivery and access. SAVANT aims to build on metadata standards, such as MPEG-7, TV-AnyTime and MPEG-21, to support the different aspects of automatic adaptation and personalisation. This paper examines different metadata standards for providing scalable content and services within the SAVANT system (Section 3) motivated by the SAVANT scenarios (Section 2).

2. SAVANT scenarios

SAVANT has identified a number of scenarios that are representative of the type of service that the project aims to develop. We describe two here: “Latest news anytime anywhere” and “Enhanced and interactive sports”. These scenarios envisage the use of four different user devices:

- *Set Top Box (STB)* with TV and remote control forms the core of the overall system. It provides large storage and is used as a gateway server to the connected user devices. Its components include a wavelancard, IP connection (also accessible via UMTS) and a DVB card.
- *Tablet PC*, operated via touch screen and keyboard, is used as a portable device at home. Its components are a DVB-T card and wavelancard for the local network.
- *PC* is a stationary terminal equipped with a media player and fast Internet access.
- *PDA* is used as the mobile device for “on the move” usage. It is equipped with storage media and can automatically connect to the STB. Its components are wavelancard and UMTS connection.

Independently of their location and device, the users of a SAVANT terminal will access a wide variety of media content, such as HTML and XML for textual data, MPEG-2 and MPEG-4 for audio/visual content, JPEG for images and MP3 for audio, in parallel or in addition to the main broadcast content.

Latest news anytime and anywhere: This service scenario describes the daily routine of a fictitious person – Mr X – who wants to be well informed and provided with personalised news. Mr X uses a multitude of different devices in a number of different contexts: STB and Tablet PC at home, PDA on the move, and PC at work. He is able to watch the intelligently recorded morning news broadcast in time-shifted view, select news items of interest and compose his own personal news programme. He has access to a number of scalable services such as personalised news summaries, enhanced high quality audio and video content, MPEG-4 videos, Web content, additional languages and signer, both via push and pull channels. He can receive personalised news alerts and use interactive services such as voting. The scenario focuses on the scalability aspects of content and services delivered to multiple devices and demonstrates the following technologies: scalable audio/visual and Internet content on STB, PC, Tablet PC and PDA, scalable services, remote user interactivity, message alerts, personalisation and dynamic user profiling, and time-shifted viewing.

Enhanced and interactive sports: This scenario considers both group and individual viewing experiences and focuses on the use of multiple devices at the same time with both synchronous and asynchronous content delivery. The viewers are able to watch the main broadcast on a TV screen while receiving additional broadcast and Web content, such as the feed of additional camera angles, audio commentary in different languages, subtitles, leader-board Internet pages, on their PDA. Several scalable services are demonstrated such as personalised highlights, betting services, authorisation of premium content, participation in a quiz, on-line shopping, 3D object modelling (e.g. a golf

course), and 3D motion analysis (e.g. tennis serve). The scenario demonstrates synchronised and scalable content and services on STB and PDA.

3. The SAVANT system

The SAVANT system has three main components, as shown in Figure 1: a) the content creation and annotation system, b) the content delivery system focusing on the synchronisation and smart delivery of content via multiple transmission channels e.g. DVB, IP, and c) the content access system, which includes sub-systems for access management, search and retrieval, synchronisation and presentation. Based on this architecture users can gain access to integrated services and multimedia content from different and combined information sources (e.g. broadcast, Internet) on a variety of user terminals.

The cornerstones of the content access system are the use of different devices with different capabilities and the personalisation of content and services delivered to these devices. These functions require the adaptation of content and services to user preferences and terminal capabilities. Personalisation includes adaptation to the user's preferred audio or display settings, personal interests and background. Adaptation to personal interests involves tasks like generation of personalised text or video summaries, highlights, recommendations and message alerts. Adaptation to the technical infrastructure available to the user is driven by technical parameters like resolution of images and frame rate of videos, but also by means of media substitutability like substituting MPEG-2 with MPEG-4 video, audio with text, text with voice synthesis, or MPEG-4 video with a sequence of pictures or

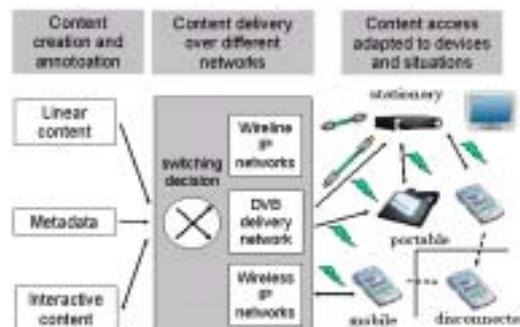


Figure 1. The SAVANT system

animation.

Adaptation can be done with or without the use of metadata. For example, it is possible to transcode MPEG-2 video to MPEG-4 based on features such as scene changes and closed captions. However, this is currently not practical due to the failure of automatic video analysis techniques to correctly identify

meaningful segments. In SAVANT we will follow the semantic approach to content adaptation. The semantic adaptation of content and services requires a metadata-based description of user preferences, terminal capabilities and valid content and services that can be adapted. With this aim we investigate several metadata standards in the next section.

4. Metadata for content and service adaptation

The scenarios, described in Section 2, focus on the adaptation of content and services to user preferences and terminal capabilities. This requires metadata at different levels. The content elements intended for adaptation at the service platform or at the terminal must be annotated with metadata at the media stream level. Semantic annotation at object level is also useful for search and retrieval based on entities such as persons, locations and times. A higher level of metadata is necessary to describe the service and the options of the service, including the references and relationships between media elements and service components. To support search, user profiling and recommendation functions a programme level metadata is required. Furthermore, metadata to describe the user preferences and terminal capabilities is also necessary. In the following we describe the metadata standards that are considered as solutions for the above requirements.

Media stream level metadata

MPEG-7 [2] is an extensive and extendible metadata standard that provides a rich set of tools to describe the structure and semantics of multimedia content. An MPEG-7 Descriptor (D) can describe both low-level features such as colour or texture characteristics, and high-level features that carry semantic meaning such as location and person names. An organised collection of Ds defines a Description Scheme (DS), which enables the description of complex objects, such as persons or events, associated with the multimedia content. The overall syntax of MPEG-7 descriptors is defined by the Description Definition Language (DDL). MPEG-7 metadata can be associated with media streams, such as MPEG-2 and MPEG-4, and can be inserted as additional information into the transport stream. At the user terminal, MPEG-7 can be used to locate structural or semantic components of a currently viewed or stored content. This facilitates search and retrieval allowing users to access parts of the data that is of interest to them [3].

Within SAVANT, MPEG-7 annotation is required to support functions like retrieval of additional information related to the currently viewed content, filtering according to user interests, generating highlights, summaries or semantically linked video chains (for example to compose a virtual feed for following a golf player). We consider the use of several Multimedia DSs for this purpose, including Segment DSs for describing the temporal and spatial

structure of the multimedia content, StructuredAnnotation and FreeTextAnnotation DSs for semantic description in a structured (Who, What, When, Where etc.) and natural language form. Using the StructuredAnnotation DS, the problem of following a golf player for example, could be solved by setting up a filter such as //AudioVisual/SegmentDecomposition/StructuredAnnotation/Who[.="Tiger Woods"]. Finally we aim to employ the Creation and Classification DSs for categorisation to support personalisation at the media stream level. We propose to modify the Classification DSs of the MPEG-7 standard in order to allow for differentiating between different news items.

Programme level metadata

TV-Anytime [4] is a metadata standard developed to define specifications for programme level content descriptions to allow viewers to find, navigate and manage content from a variety of sources, including enhanced broadcast, Internet and local storage. Such metadata includes attractors (title, synopsis, genre, cast and awards etc.) that aid the acquisition of available content organised in EPGs (Electronic Programme Guide).

TV-Anytime metadata will be used in SAVANT to provide a high-level programme description, which is necessary to enable filtering, recommendation, and search and retrieval of content from the programme listings of EPGs. We consider the adoption of the following elements of the ProgramInformation DS: BasicContentDescription DS for programme description and classification, AVAttributes DS for technical description. We propose to modify the current classification schema of TV-Anytime by separating certain categorisation classes, such as target audience, into orthogonal classifications and by extending the Genre DS.

User preferences and terminal capabilities

Recently the MPEG forum started work on a new standard, MPEG-21 [5], with the aim of defining a framework to enable transparent access to multimedia resources across a wide range of networks and devices. Toward this goal, MPEG-21 targets the adaptation of Digital Items, which are defined as structured digital objects with standard representations, identifications and descriptions. MPEG-21 describes a variety of dimensions, including terminal, network, delivery, user and natural environment capabilities. Terminal capabilities include hardware properties such as processor speed, software properties such as operating system, display properties such as screen resolution, and device profiles indicating the supported media formats (e.g. MPEG-2). Network capabilities specify delay, error and bandwidth characteristics. Delivery capabilities specify the types of transport protocols supported (e.g. TCP/IP) and the types of connections (e.g. multicast). User preferences include display, accessibility and mobility characteristics and natural environment characteristics include location related information.

Due to the current working status of the MPEG-21 standard we also considered the use of the CC/PP (Composite Capability/Preference Profiles) standard [6], developed by the Web and mobile phones community. A CC/PP profile is a description of device capabilities and user preferences consisting of a number of components (client, proxy), each containing a number of attribute names and associated values that are used by a server to determine the most appropriate form of a resource to deliver to a client.

MPEG-21 or CC/PP will be used in SAVANT to maintain descriptions of terminals and networks to drive adaptation based on these preferences. Adaptation according to user preferences will be based on MPEG-7, which also forms the basis of the MPEG-21 user preferences schema. To support user profiling based on a probabilistic framework, and to allow for the modelling of different user interests with respect to different news items we propose to modify the UserPreferences DSs of the MPEG-7 standard.

How will they work together?

Content and services will be annotated with MPEG-7 at the media stream level and TV-Anytime at the programme level by the content/service provider using semi-automatic annotation tools. The main broadcast content (MPEG-2) and any additional content (MPEG-4, HTML etc.) combined with the MPEG-7 and TV-AnyTime metadata will be treated as Digital Items and annotated with MPEG-21 for delivery over different networks and to different terminals. User terminals will maintain descriptions of user preferences (MPEG-7) and terminal capabilities (CC/PP or MPEG-21), and perform content and service adaptation tasks. The STB acting as a gateway will have additional functions such as maintaining communication with connected mobile devices, perform transcoding as a proxy (e.g. for PDA), and gather and manage annotation data.

5. Conclusions

In SAVANT, we see the combination of metadata standards, in particular MPEG-7, TV-AnyTime and MPEG-21, as a solution for describing scalable broadcast and Internet content and services, and driving their adaptation according to different user preferences and terminal capabilities. Future work in SAVANT will focus on the design and implementation of the SAVANT system to provide annotation, delivery and access of multimedia content according to the combined use of these standards.

References

1. SAVANT, <http://www.ist-savant.org>
2. ISO MPEG-7, Part 5 - Multimedia Description Schemes, ISO/IEC JTC1/SC29/WG11/N4242, (Oct 2001)

3. A. Pearmain, M. Lalmas, E. Moutogianni, D. Papworth, P. Healey and T. Roelleke. Using MPEG7 at the Consumer Terminal in Broadcasting, *EURASIP, Journal on Applied Signal Processing*, (2002)
4. TV-Anytime Provisional Specification 1.3, (Sept 2002)
5. ISO MPEG-21, Part 7 - Digital Item Adaptation, ISO/IEC JTC1/SC29/WG11/N5231, (Oct 2002)
6. Composite Capability/Preference Profiles, W3C Working Draft, (March 2001)