A metadata model supporting scalable interactive TV services

Gwenaël Durand
EXPWAY
Paris, France
gwenael.durand@expway.com

Gabriella Kazai, Mounia Lalmas
Queen Mary University of London
London, UK
{gabs, mounia}@dcs.qmul.ac.uk

Uwe Rauschenbach
Siemens AG, Corporate Technology
Munich, Germany
uwe.rauschenbach@siemens.com

Patrick Wolf
Fraunhofer IPSI
Darmstadt, Germany
Patrick.Wolf@ipsi.fraunhofer.de

Abstract

In this paper, we introduce a novel metadata model for describing scalable and interactive TV services that can be enriched with supplemental multimedia information. The model allows users to access such TV services not only via their traditional TV sets, but also via additional mobile devices like TabletPCs or PDAs. To achieve this, we segment the traditional linear program into sub-components, while separating device-independent and device-specific metadata. A realization of this model builds on the existing standards of TV-Anytime, MPEG-7 and MPEG-21. The model achieves a step towards the "Connected world" vision of the second specification phase of the TV-Anytime forum.

1. Introduction

Today, both multimedia computing and television provide entertainment to a broad audience. While television efficiently transmits the same content to many viewers at a time, multimedia computing provides a highly interactive and personalized experience. The advent of digital television allows the combining of both these worlds [7, 6]. In a digital television system, additional media content and interactive applications can be broadcast, allowing the viewer to become involved and to change the classic linear program flow by interacting where desired. Complementing the TV set, additional mobile devices can be used to consume the personalized content. For example, a TabletPC with its large screen provides for interactive portable in-home use, and a PDA allows convenient mobile access to the content both in-home and on the move. As these devices have different capabilities in terms of, e.g., screen space and decoding power, scalability of the provided service becomes a crucial issue [9].

Creating interactive TV programs using these new opportunities poses new challenges. Content needs to be created in a suitable format or adapted to the different devices and user requirements. The service, consisting of such contents, needs to be scaled accordingly. To allow this, the service must be complemented by a consistent set of metadata, which describes the additional content options, the possible interactions and the scalability parameters. Such a service description has to be based on open, established standards, allowing a large variety of current and future terminal devices to consume the new services.

In the European IST SAVANT project [1], convergence concepts for scalable interactive TV considering multiple terminal device types and different transmission networks are investigated along with metadata frameworks to support the realization of such concepts. This paper describes the metadata model developed as part of this activity.

The paper is structured as follows. First, the concept of a scalable interactive TV service is introduced. Second, an abstract data model is presented, which has the power to describe such a service. Third, the foundations for realizing our metadata model based on established multimedia metadata languages are described, and the approach taken to realize the model is presented. We close with conclusions and future work.

2. Concept of a scalable interactive TV service

The concept of a scalable interactive TV service covers a broad range of aspects from content generation to content access. Most prominently, such a service aims to create and deliver enhanced and interactive content tailored to different user devices, varying network conditions and diverse

user requirements. To support this notion of scalability, we consider a service as a collection of service components, where a service component is either a segment of the *main broadcast content* or an item of the *additional related multimedia content*. In a news broadcast program, for example, the individual news items (news stories) provide the main content, where each item corresponds to an individual service component. Additional content, i.e. content provided in the realm of digital TV to enrich a program, may include bonus MPEG-4 video clips, HTML pages, 3D graphics, additional languages (both audio and subtitles), a signer, interactive games, and so on.

The breakdown of a service into service components supports its adaptation for a given device type, network or user by allowing the sorting, filtering and adaptation of its individual components. A service as such may also be dynamically modified according to events, such as the availability of newly produced content. Different service components can be routed over different transmission networks (e.g. DVB, IP) and be combined at the receiver side. In addition, the content of a scalable service can be accessed not only from a single device, but also from a combination of devices available to users. For example, while watching the live broadcast of a football match on the TV screen, up-to-date game statistics (e.g. goals, fouls, etc.) may be viewed in parallel on a PDA whenever requested by the viewer.

The adaptation of a service and its components requires metadata that describe the format and semantics of each item. Based on such metadata, content delivery and access systems can make intelligent decisions as to when and how to transmit, display, adapt or ignore a service component. This metadata, defined within an appropriate schema, is referred to as a service description. Such a description, in addition to supporting scalability, also provides the necessary mechanisms to structure service components within the service (e.g. relate and synchronize additional content items with the main content).

In order to be applicable in the context of a TV program, the service description must be an extension of the conventional linear TV content model. Furthermore, as it is likely that interactive TV services will evolve dynamically, the metadata schema has to be extendable in order to cater for new service concepts. Content semantics and structure must be clearly separated from device-specific realization details. The following section proposes a metadata model for a service description, which meets these requirements.

3. A metadata model to support scalable TV services

This section details the conceptual model that provides the basis for our metadata framework for describing scalable interactive TV services. The model, outlined in Figure 1, describes such a service as an enriched TV program. We detail this model by introducing the underlying rationale that led to its development.

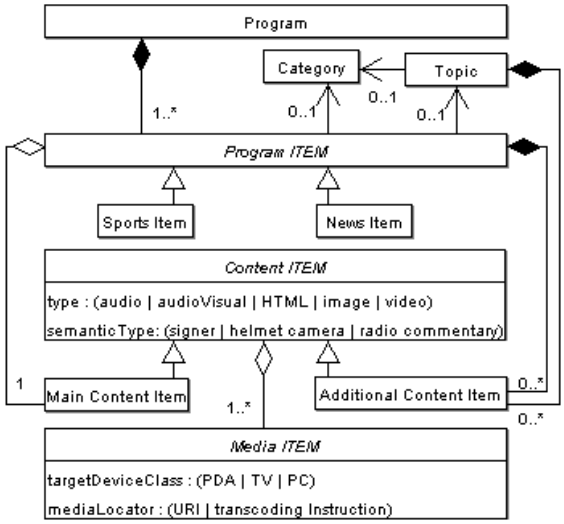

Figure 1. Overview of the metadata model

***Program* as anchor to traditional TV.** In the broadcast world the concept of a program is well established: it "represents an editorially coherent piece of content" [4]. In order to comply with this traditional usage, it is essential that a metadata model does not change, but only extends the notion of a linear program. To reflect this, the *Program* element has been introduced as the root of our model. It acts as the anchor to the traditional TV world, while it also serves as the basis for our extensions.

Segmentation into *Program Items*. As mentioned in the previous section, we define a program as a collection of service components. These are represented by a sequence of *Program Item* elements, segmenting the linear TV content. Each *Program Item* is considered as a coherent and semantically closed entity, such as a single news story (with its main and additional content). *Program Items* act as containers for both descriptive elements, including title, keywords and *Category*, and structural elements like the media content associated with a *Program Item*.

Unlike other metadata models, such as the digital broadcast item model [11], which allow the decomposition of a linear broadcast on the structural level, the model presented in this paper supports decomposition on a semantic level. *Program Items* allow the representation of semantic units designed and created by editors with a specific purpose or meaning within the context of a given *Program*. By explicitly supporting the representation of such semantic units, the

proposed model allows content creators to directly think in semantic contexts.

Device-independent content descriptions. The media content of a *Program Item* contains one segment of the linear broadcast content combined with a rich and diverse, but optional, set of additional content. To support the rendering of such content on various terminal devices, multiple device-specific formats of the same content may be included within the broadcast. In order to allow for the description of the *same* content for various devices, we introduce the notion of *Content Items* in our model. *Content Items* provide a device-independent abstraction layer for describing only the *content* of the media without being specific regarding its concrete physical representation. At this level, aspects that influence how users interact with the content such as its *Type* (e.g. HTML page, audio track, video) and *SemanticType* (e.g. signer, helmet camera) are considered, while the technical details, i.e. the exact format (e.g. MPEG-2, MPEG-4), are left open. We distinguish two types of *Content Items*: *Main Content Items* (MCI) and *Additional Content Items* (ACI). MCIs describe the linear broadcast segments that make up the basic service and ACIs describe the supplemental information. For example, a news story in a news show that is enriched with a signer and supplemented with background information from the web would be described as a *Program Item* with three *Content Items*: one MCI and two ACIs (the signer as an ACI of type "video" and the web pages as an ACI of type "HTML").

Device-specific technical details. *Content Items* do not describe how the content is rendered at the terminal device. For instance, the signer ACI in the above example only specifies that it is a video clip to be synchronized with the main video (the MCI), but says nothing about its exact format. It may contain a clip in MPEG-2 format to be displayed on a TV or one with lower resolution in MPEG-4 for a PDA. In our model, each concrete realization of a *Content Item* is described as a *Media Item*. These correspond to separate copies of the same content, encoded in different formats or with different parameters (e.g., resolution, bit rate). A *Media Item* has a *targetDeviceClass* that describes the device class for which the content was designed (e.g. TV, PDA). A device class maps directly onto basic technical parameters such as *resolution*, *bitrate*. Each *Media Item* contains a *mediaLocator*, which is a URI that either points to an actual, already existing essence or contains a transcoding instruction. In order to render a *Content Item* at a terminal device, the *Media Item* that best matches the terminal type is chosen, and the appropriate player (i.e. MPEG-4 player or HTML browser) is launched on the terminal.

Beyond *Program Items*: Besides supporting the inclusion of *Additional Content Items* at the *Program Item* level, it is often desirable to attach additional content to an entity which is not as fine-grained as a *Program Item*. This entity is provided by the *Topic* element. A *Topic* provides a "bigger story", grouping several related *Program Items* (e.g., "Presidential election" in a news program or "Athletics world cup 2003" in a sports program). *Topics* can be described and classified; and additional content may be provided related to *Topics*. Typically, a *Topic* has a lifetime of a few days to a few weeks, during which new *Program Items* and new additional content may be added.

4. Realization of the metadata model

In order to be a part of the broadcast value chain, the proposed model has to be transformed into a concrete service description. This service description has to be based on existing standards to allow interoperability of the many components from different vendors found in a broadcast system. This section analyzes the suitability of existing metadata standards to reflect different aspects of the proposed model, provides the rationale behind selecting one of them to base the service description upon and describes the necessary extensions to fully describe the proposed model.

4.1. Existing metadata standards

Based on previous studies [12, 8, 10, 13], three established standards, MPEG-7 [2], TV-Anytime [4] and MPEG-21 [3] were selected to provide the basis for describing scalable TV services. The strength of MPEG-7 is the description of multimedia content. MPEG-21 provides (among others) structuring tools for digital items and descriptors for network and device capabilities. TV-Anytime describes TV services, mainly for content selection and time-shifted viewing.

To realize a service description for scalable interactive TV, the requirements, summarized in Table 1, must be met by the metadata language to be employed.

Since no single metadata language meets all requirements perfectly, the service description is realized by combining elements from all three standards. TV-Anytime is used to provide the overall framework, as its structure is best suited to describe the semantics of a TV service. The extensibility mechanism of TV-Anytime (replacing elements by those of derived types) is exploited to provide the additional elements as needed. As TV-Anytime allows the use of MPEG-7 elements for describing, e.g., media formats, the advantages of MPEG-7 can be exploited in the leaf nodes of the service description, while the higher level nodes are taken from TV-Anytime. MPEG-21 elements can be included by deriving new types from existing TV-Anytime types and including the MPEG-21 nodes into these types. This way, the extended description remains compatible with TV-Anytime, hence allowing a TV-Anytime-compliant Set Top Box to use the TVA-compliant features.

Requirement	MPEG-7	MPEG-21	TV-Anytime
Ability to describe interactive TV services	-	o	+
Express hierarchy of service description model	o	+	o
Easy interpretation of the semantics of an element	+	-	+
Standardized, extendable metadata language	+	+	+
Pre-defined constructs to describe elements of service description model	+	+	-

Table 1. Compliance of metadata languages to requirements (+ supported, - not supported, o: partly supported

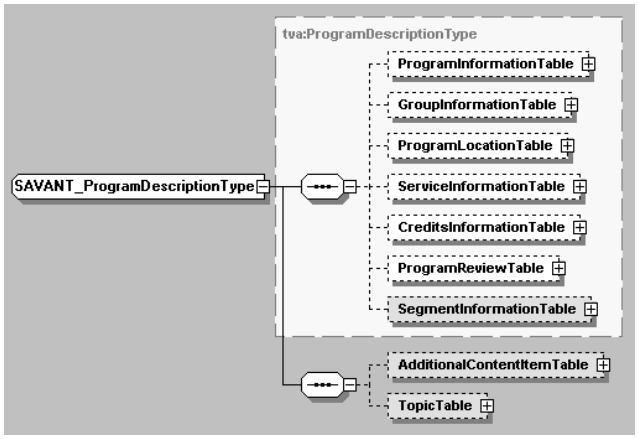

Figure 2. Extended TV-Anytime program description

4.2. Extension elements

TV-Anytime provides tools for describing sets of TV programs and their schedule over one or more TV channels. Our model extends these tools in order to significantly enrich the description of each individual program, while also supporting a semantic description level and the description of different, scaled, versions of the same content items.

Figure 2 illustrates the extensions applied to the TV-Anytime model. A TV-Anytime description is composed of a set of top-level tables containing different elements of the description. Original tables are contained in the yellow box. Most of our extensions are contained in the 2 additional tables and within the extended *SegmentInformationTable*.

SegmentInformationTable: Segment information elements are used in TV-Anytime for defining temporal units of individual programs. This then allows to describe a summary of the content with highlights or to bookmark parts of a program. In our extension of TV-Anytime, we mapped the individual *Program Items* of our metadata model onto these *Segment Information* elements. In order to support the *Program Item* structure (as defined in section 3), we extended the original TV-Anytime elements, whereby each extended segment information contains at least one *Media Item* element, can be linked to a *Topic* element, and can have an extended MPEG-7 *summary*. The segmentation of a program into segments may follow structural or semantic decomposition. The description of a *Program Item* at the semantic level is supported by the extended *BasicSegmentDescriptionType* of TV-Anytime. The new *Media Item* element describes a device-specific instance of the semantic unit of a segment. It contains elements based on the MPEG-21 standard, such as *TargetDeviceClass*, *FileFormat* and *MediaLocator*.

AdditionalContenItemTable: The *Additional Content Items*, enriching a digital TV program, are described in this newly-introduced table. Each entry describes an ACI, where the description of an ACI element contains:

- References to the *Topic* and the list of *Program Items* that the ACI is related to;
- *Synchronization information* defining how the ACI should be synchronized with the MCI. This is controlled by two parameters: *MediaTime* and *Tolerance*. The former builds on standard MPEG-7 descriptors, e.g. *MediaRelIncrTimePoint* and *MediaIncrDuration*, while the latter is a new parameter describing the accepted level of delay in content delivery;
- *Routing information* providing hints to the system on how the ACI instance should be delivered to the terminal (e.g. over broadcast or IP channel);
- A set of *Media Items*, each describing an actual device-specific instance of the ACI.

TopicTable: elements of this table define *Topics* that allow the grouping of *Program Items* and *Additional Content Items*.

The example XML fragment in Figure 3 illustrates the basic structure of our metadata model along with some examples of device-specific *Media Items* of a given segment and an example ACI. Note that for simplicity, some attributes and elements have been omitted.

```
<TVAMain xmlns="urn:svt:metadata" ...>
 <ProgramDescription>
  ...
  <SegmentInformationTable>
   <SegmentList>
    <SegmentInformation segmentId="..."
      fragmentId="89" fragmentVersion="0"
      topicId="EasyJET">
     <ProgramRef crid="crid://..."/>
     <Description>
      <Title>...</Title>
      <Synopsis>...</Synopsis>
      <Keyword>...</Keyword>
     </Description>
     <SegmentLocator>
      <mpeg7:MediaRelIncrTimePoint>...
       </mpeg7:MediaRelIncrTimePoint>
      ...
     </SegmentLocator>
     <svt:MediaItem>
      <svt:TargetDeviceClass>STB
       </svt:TargetDeviceClass>
      <FileFormat>mp2</FileFormat>
      <FileSize>...</FileSize>
      <BitRate>...</BitRate>
      <svt:MediaLocator>
       <mpeg7:MediaUri>...</mpeg7:MediaUri>
      </svt:MediaLocator>
     </svt:MediaItem>
     <svt:MediaItem>
      <svt:TargetDeviceClass>PDA
       </svt:TargetDeviceClass>
      <FileFormat>mp4</FileFormat>
      ...
     </svt:MediaItem>
     <svt:Summarization>
      <mpeg7:Summary>...</mpeg7:Summary>
     </svt:Summarization>
    </SegmentInformation>
    ...
   </SegmentList>
  </SegmentInformationTable>
  <svt:AdditionalContentItemTable>
   <svt:AdditionalContentItem aciId="..."
     type="audio" topicId="TourDeFrance"
     preferredRouting="DSLStreamIP" ...>
    <svt:Description>...</svt:Description>
    <svt:MediaItem>...</svt:MediaItem>
    <svt:SyncInfo>
     <svt:MediaTime>...</svt:MediaTime>
     <svt:Tolerance>medium</svt:Tolerance>
    </svt:SyncInfo>
   </svt:AdditionalContentItem>
   ...
  <svt:AdditionalContentItemTable>
  ...
 </ProgramDescription>
</TVAMain>
```

Figure 3. A sample XML instance of the SAVANT metadata model

4.3. Fragmentation mechanisms

The proposed metadata framework also includes mechanisms to suitably fragment the metadata such that the fragments can be individually transmitted and accessed, saving transmission bandwidth while still allowing to tune into the program at any-time.

To enable the efficient delivery, updating and navigation of a metadata description, TV-Anytime defined a generic decomposition mechanism of the description into self-consistent units of data, called TVA fragments. In our framework, we extended the set of normative TVA fragments in order to include AdditionalContentItem fragments and Topic fragments.

TV-Anytime's fragment identification and versioning mechanisms allow *individual* fragments to be transmitted, udpated, accessed and processed on the end-user terminal. Finally, the TVA's fragment encapsulation mechanism was used in order to group related fragments together in to larger so-called containers for transmission.

5. Conclusions and Future Work

In this paper, we reported on a novel metadata model for describing scalable interactive TV services that are modeled as sets of inter-related service components. For each component, a clear separation of semantics and device dependent metadata has been achieved. A realization of this model by extending the TV-Anytime standard has been described. Based on this, a prototype of an interactive TV news service has been implemented and successfully demonstrated at the "Internationale Funkausstellung 2003" broadcast fair in Berlin and a combined interactive news and sports service at the "International Broadcast Fair 2004" in Amsterdam. Both these services allowed personalized and interactive access to both main content (news stories and sports content) and additional content, using a TV set, a TabletPC and a PDA.

We believe that the metadata model presented here is a first step towards the achievement of the "Connected world" vision as put forward in the second specification phase of the TV-Anytime forum [5]. Beyond traditional TV, this vision aims to address all kinds of connected devices, non-linear program structures and new content types.

6. Acknowledgement

SAVANT is funded by the European Union under the contract number IST-2001-34814. We thank all members of the SAVANT consortium, especially John Cosmas, Wolfgang Putz, and Jürgen Lang for their input to this work.

References

[1] SAVANT project home page. http://www.savant.tv.

[2] ISO MPEG-7, part 5 - multimedia description schemes. ISO/IEC JTC1/SC29/WG11/N4242, 2001.

[3] ISO MPEG-21, part 7 - digital item adaptation. ISO/IEC JTC1/SC29/WG11/N5231, 2002.

[4] TV-Anytime phase 1, part 3 metadata, sub-part 1: Metadata schemas. ETSI TS 102 822-3-1, 2002.

[5] Requirements series, R-1 (phase 2) on business models. ftp://tva:tva@ftp.bbc.co.uk/pub/Plenary/TV140r8.zip, 2003.

[6] G. Bell and J. Gemmell. A call for the home media network. *Communications of the ACM*, 45(7):71–75, July 2002.

[7] Y. Gonno, F. Nishio, T. Tsunoda, and Y. Yamagishi. White paper on integrated broadband environment for personalized TV experience (IBEX). In *Proc. ACM Multimedia*, pages 63–66, Los Angeles, 2000.

[8] G. Kazai, M. Lalmas, M.-L. Bourguet, and A. Pearmain. Using metadata to provide scalable broadcast and internet content and services. In *Proc. WIAMIS*, pages 487–492, London, April 2003.

[9] A. Lugmayr, J. Lyytinen, A. Mailaparampil, F. Tico, S. Maijala, S. Tuominen, T. Pihlajamaeki, P. Rautavirta, M. Oksanen, J. Spieker, S. Niiranen, and S. Kalli. The future interaction TV project developing diet - digital interaction environment for TV. In *Proceedings of NORSIG*, Norway, October 2002.

[10] A. Lugmayr, S. Niiranen, and S. Kalli. Harmonizing broadcast metadata standards - the digital broadcast item model. In *Proc. WIAMIS*, pages 478–482, London, April 2003.

[11] A. R. Lugmayr, A. B. Touimi, I. Kaneko, J.-N. Kim, C. Alberti, S. Yona, J. Kim, M. T. Andrade, and S. Kalli. MPEG-21 in broadcasting: the novel digital broadcast item model. In *Proceedings of SPIE, Real-Time Imaging VIII*, volume 5297, pages 163–174, May 2004.

[12] S. Pfeiffer and U. Srinivasan. TV-Anytime as an application scenario for MPEG-7. In *Proc. ACM Multimedia*, pages 89–92, Los Angeles, 2000.

[13] C. Tsinaraki, E. Fatourou, and S. Christodoulakis. An ontology-driven framework for the management of semantic metadata describing audiovisual information. In *Proceedings of the 15th Conference on Advanced Information Systems Engineering (CAiSE '03)*, volume 2681/2003 of *Lecture Notes in Computer Science*, pages 340–356, Klagenfurt/Velden, Austria, 2003. Springer.