Metadata-Based Content Management and Sharing System for Improved User Experience

Markus Waltl, Christian Raffelsberger, Christian Timmerer, and Hermann Hellwagner

Klagenfurt University, Inst. of Information Technology, Multimedia Communication Group
Universitätsstraße 65-67, 9020 Klagenfurt, Austria
{firstname.lastname}@itec.uni-klu.ac.at

Abstract. In the past years the amount of multimedia content on the Internet or in home networks has been drastically increasing. Instead of buying traditional media (such as CDs or DVDs) users tend to buy online media. This leads to the difficulty of managing the content (e.g., movies, images). A vast amount of tools for content management exists but they are mainly focusing on one type of content (e.g., only images). Furthermore, most of the available tools are not configurable to the user's preferences and cannot be accessed by different devices (e.g., TV, computer, mobile phone) in the home network. In this paper we present a UPnP A/V-based system for managing and sharing audio/visual content in home environments which is configurable to the user's preferences. Furthermore, the paper depicts how this system can be used to improve the user experience by using MPEG-V.

Keywords: Metadata, UPnP A/V, Content Management, Content Sharing, MPEG-V, Sensory Effects, User Experience.

1 Introduction

The amount of content, i.e., movies, images, and music, has drastically increased in the last years [1]. As a consequence, users have increasing difficulties in managing their content and finding a specific movie or image. There are different techniques (e.g., searching, browsing) to reduce the amount of data a user has to look through before finding the desired content. The major problem with the default techniques is that they are not configurable to the user's preferences and to the content itself.

The issue of not being able to configure the search/browse engine to the user's preference is addressed by the metadata-based *Content Management and Sharing System* (CMSS) presented in this paper. The CMSS provides the ability to browse content in a configurable tree instead of searching for content. This approach provides the user the functionality to browse for a specific content (e.g., movies from the action genre released during the 1980s). Furthermore, the browsing tree provided by the CMSS can be dynamically changed through a user-editable configuration. This allows the user to adjust the system to her/his preferences. Furthermore, the configuration can support all kinds of content, e.g., music, images, videos, or MPEG-V metadata.

F. Alvarez and C. Costa (Eds.): UCMEDIA 2010, LNICST 60, pp. 132–140, 2012.

The support of all kinds of content allows for a wide range of possible usage scenarios, e.g., for increasing the user experience during consumption of multimedia content. Moreover, the CMSS allows editing of metadata (e.g., author, title) that belongs to specific content and provides content via UPnP A/V [2] and RTP/RTSP [3,4]. The content offered by the CMSS can be consumed by a large number of devices without special requirements (e.g., a set-top box) because of the usage of the UPnP A/V standard.

The remainder of this paper is organized as follows. Section 2 describes the architecture of the CMSS which has been developed during the course of the INTERMEDIA project. Section 3 provides an overview of MPEG-V Media Context and Control which is a standard under development for enhancing the user experience. Section 4 presents how to combine it with the CMSS. Conclusions and future work items are presented in Section 5.

2 Architecture of the Content Management and Sharing System

Fig. 1 depicts the architecture of the CMSS. Most of the CMSS is based on Java for providing portability to various systems, only some external libraries are not written in Java (e.g., native extraction libraries). The CMSS consists of components for managing, storing, importing and streaming multimedia content. The system is a Web application running on a JBoss Application Server [5] and uses MySQL [6] as metadata storage. However, the CMSS comprises two different types of storages: one for storing metadata using MySQL and one for storing the actual media content utilizing the file system. The CMSS provides a Web interface to enable a user to upload new content to the system or to modify existing metadata. On the other side of the chain, the user can browse this content via UPnP A/V [2] which is a well established standard supported by many multimedia devices and applications. UPnP A/V provides services for retrieving a list of the available contents and for requesting a specific content. The user at the consumer side can use UPnP A/V for retrieving the desired content which will be transferred via HTTP to the MediaRenderer or streamed via RTP/RTSP [3,4] to a media player. When using RTP/RTSP, the user has a control mechanism over the data stream of the content. She/he can play, pause, rewind, or fast forward the content without downloading the whole content like in HTTP. The RTP/RTSP support is provided by the *VideoLan Manager* (VLM) that is a component of the *VideoLan Player* (VLC) [7]. For providing UPnP A/V within the CMSS, the *Cyberlink for Java* [8] library is used. To be compatible with UPnP A/V, software applications or hardware devices have to include at least one UPnP A/V device (MediaServer, ControlPoint or MediaRenderer) and its mandatory services. The CMSS includes a UPnP A/V MediaServer. There are applications providing several devices, thus, the logical separation may not be found in practice (e.g., a TV that implements both ControlPoint and MediaRenderer).

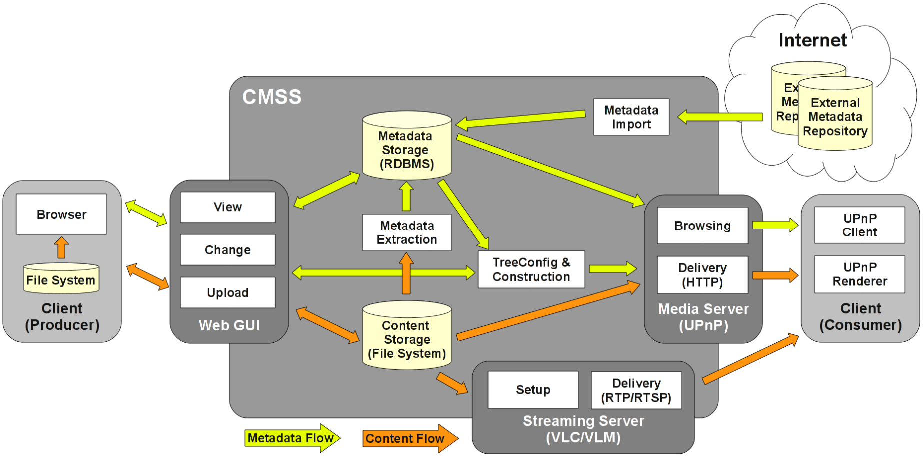

Fig. 1. Architecture of the Content Management and Sharing System

The CMSS offers various tools that reduce the effort for the user to add metadata to newly uploaded content. For example, low-level metadata (i.e., bit rate, codec, resolution, etc.) is automatically extracted by an extraction chain consisting of *Mime-util* [9], *Libextractor* [10] and *FFmpeg* [11]. These tools can be replaced or extended with additional extractors. The resulting low-level metadata information is stored within the metadata storage for later retrieval. The content itself is placed in the content storage.

After uploading, the user has the possibility to add high-level metadata (i.e., cast, director, artist, plot, etc.). This can happen manually or via metadata importers. Currently, two importers are available: for videos, from the TheMovieDB.org [12] repository and, for music, from the MusicBrainz [13] repository. The CMSS supports adding new importers as well as modifying and removing the existing importers.

A problem which conflicts with the user-centric approach is the pre-defined (static) structure of the content list in many UPnP A/V MediaServers. Therefore, the CMSS allows the user to customize the view (i.e., presented tree structure) on the metadata via the Web interface. The configuration of the view is stored in a text file which allows the user to customize the tree structure in an easy way. We selected a simple text based structure instead of other more complex representation forms (e.g., XML) because it is easy to learn, read and modify. Note that the system can be modified to allow other configuration representations. The configuration consists of one or more configuration lines, where each configuration line consists of one or more keywords. These keywords are mapped to certain metadata tags that are included in the database (e.g., title, genre, resolution, language). Additionally, two special keywords (*Static* and *Alphabet*) are defined. The *Static* keyword is used for providing a label which can be used for renaming or restructuring the content tree. The *Alphabet* keyword can be used to further improve the structure by automatically grouping a long list of nodes (e.g., performers, titles) based on their initial character. We limited the *Alphabet* keyword to only support a subset of the available database fields because it is not very useful to define *Alphabet* for various fields (e.g., numerical fields such as bit rate, resolution, duration). Each line in the configuration file denotes a branch in the tree. To build a hierarchy of nodes, multiple nodes can be linked with the "->" operator. The node at the left side of the operator is the parent, the right one is the child. Equal, or partially equal, branches are automatically merged.

This functionality is an approach to provide a user-centric content presentation which allows the user to personalize the view on the content. For example, if the user only wants to have images and music, she/he can remove the entries for videos from the configuration.

An example for a small CMSS tree configuration is shown in Listing 1 and an abstract representation of the resulting tree is depicted in Listing 2. The example presents a tree which consists of a static label *Genres* that is defined by *Static:Genres*. The next keyword *Genre* specifies to query all genres that exist in the database and to insert them into the tree. Every genre is further divided into two separate branches: *Performers* and *Decades*. The *Performers* label consists of a list of the names of the performers within this specific genre. The second line of the configuration includes a *Decades* label which introduces a list of decades that are based on the release date of the content. For example, a historical movie released in the year 1965 will be listed under the decade "1960s" which is located in the "History" branch.

Listing 1. Example CMSS tree configuration

```
Static:Genres->Genre->Static:Performers->Performer
Static:Genres->Genre->Static:Decades->Decade
```

Listing 2. Abstract representation of the example tree

```
Genres
  |-> Action
  |-> Horror
    |-> Decades
      |-> 1970s
      |-> . . .
    |-> Performers
      |-> Christopher Lee
      |-> . . .
  |-> . . .
```

Note that this example tree currently only comprises movies but can consist of all types of content. Furthermore, the presented configuration file of the CMSS allows, in contrast to other systems, an on-the-fly modification of the content view and could be extended in the future to provide every user a different view on the content.

During the development of the CMSS, an extension to UPnP A/V DIDL-Lite [14] was introduced. We call this extension *bundling*. The bundling feature allows the user to associate several content and metadata files for combined usage. Thus, additional information can be delivered along with the content. With this method it is possible to send external metadata files, such as subtitles, MPEG-V sensory effects, karaoke lyrics, etc. to enhance the user experience while consuming multimedia content. Bundling information is stored within the `desc` element that is already defined in DIDL-Lite. Note that the `desc` element requires a namespace that defers from the DIDL-Lite namespace for all child elements. Therefore, we redefined the existing `res` element within a new namespace. These elements contain all information that is needed to retrieve a bundled item. For example, if a movie contains only an English subtitle but the user wants to have other subtitles she/he could bundle them to the movie (e.g., German or French subtitles) which are available as separate files. An example of such a DIDL-Lite description can be found in Listing 3. The example shows a video item that comprises information about title, artists, genres and director. Furthermore, the DIDL-Lite includes information about the location of the video. In this example, the video is available for HTTP download and as an RTP/RTSP stream. The description element announces a bundled item. Here, the bundled item represents a French subtitle file for that video.

Listing 3. Example DIDL-Lite description with bundled content

```
<DIDL-Lite>
 <item id="3" parentID="-32232" restricted="1">
  <upnp:class>object.item.videoItem</upnp:class>
  <dc:title>Babylon A.D.</dc:title>
  <upnp:artist>G. Depardieu, V. Diesel</upnp:artist>
  <upnp:genre>Thriller, Action, Sci-Fi</upnp:genre>
  <upnp:director>Mathieu Kassovitz</upnp:director>
  <res protocolInfo="http-get:*:video/mpeg:*" nrAudioChannels="5">
   http://<ip>:<port>/Export?id=3&amp;file=BabylonAD.avi</res>
  <res protocolInfo="rtsp-rtp-udp:*:video/mpeg:*"
   nrAudioChannels="5">rtsp://<ip>:<port>/vid3.avi</res>
  <desc nameSpace="urn:schemas-uniklu:09-2009:intermedia" id="3">
   <didl-ext:bundle>
    <didl-ext:res protocolInfo="http-get:*:text/plain:*">
     http://<ip>:<port>/Export?id=29&amp;file=BabylonAD_French.srt
    </didl-ext:res>
   </didl-ext:bundle>
  </desc>
 </item>
</DIDL-Lite>
```

Basically, container formats provide a very similar functionality as bundling. However, there are two main reasons why we introduced bundling instead of using existing container formats (e.g., MP4). First, a multimedia player has to be able to support the container format in order to be able to obtain the contained metadata or additional resources. Second, the whole file including all resources and metadata has to be stored and delivered to the user. There is no way to select only the desired parts of the package.

As the CMSS uses UPnP A/V to share its data with other devices, bundling should be compatible with this technology. Bundling information has to be propagated within the UPnP A/V network without breaking the compatibility with standard UPnP A/V actions. Players which do not support bundling only retrieve the main content (e.g., the video) and ignore additional information advertised in further items. Players supporting bundling can download and use the additional information, e.g., the additional French subtitles, as shown in Listing 3.

By using the bundling approach the user can freely decide what she/he wants to consume and does not have to download unwanted or unnecessary information. This reduces the needed bandwidth and storage space at the client.

3 MPEG-V Media Context and Control

MPEG-V Media Context and Control is an ISO/IEC standard under development. It will allow the integration of virtual worlds into the real world (e.g., via augmented reality) and vice versa (e.g., via avatars). For example, with appropriate devices (e.g., fans, vibration chairs, lamps) a movie can be enriched with additional effects (e.g., wind, vibration, light) for enhancing the user experience. The additional effects are described and stored in so-called *Sensory Effect Metadata* (SEM) descriptions. These descriptions are defined in Part 3 of MPEG-V [15].

The SEM description is shipped with the content (e.g., movie, audio, image) or can be downloaded from a remote storage (e.g., the Internet). Both, the SEM description and the content are delivered to MPEG-V-capable devices for playback. Fig. 2 depicts the concept of such sensory effects.

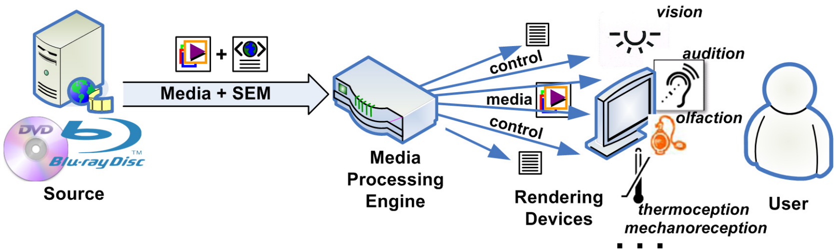

Fig. 2. The Concept of Sensory Effects [18]

Initial user studies [16,17] were conducted in the area of sensory effects for videos and yielded promising results. These results show that sensory effects are most likely to enhance the experience of users watching movies accompanied by a SEM description.

The first user study [16] presents the benefits of using sensory effects along with multimedia content. Here the participants were asked to rate the enhancement of a video with sensory effects with respect to the same video sequence without sensory effects. The results shown in [16] indicate that for different genres sensory effects are perceived differently. For example, sensory effects lead to a worthwhile user experience for action movies and documentaries while being less informative for news.

The second user study [17] presents the utility of sensory effects which means how sensory effects improve the perceived video quality. In this study two video sequences were shown with various bit rates. The videos were presented twice to the participants, one time without sensory effects and another time with sensory effects. The participants had to rate the video quality of the test sequences. The results show an improvement of the perceived video quality by using sensory effects. On average, the result for a sequence accompanied by sensory effects is 0.5 mean opinion score (MOS) higher than for the same sequence without sensory effects. More detailed results can be found in [17].

4 Usage of MPEG-V Context and Control with the CMSS

The test-bed introduced in [18] is able to work with the CMSS. The *Sensory Effect Video Annotation Tool* (SEVino) can be used for generating SEM descriptions for a video. The *Sensory Effect Simulator* (SESim) is a simulator for SEM descriptions and does not provide real-world stimulation (e.g., real wind or vibration effects). For providing real-world stimulation the *Sensory Effect Media Player* (SEMP) [16] was developed which renders sensory effects on the amBX system [19].

These tools and the CMSS are combined in the following way: on the provider side of the CMSS, SEVino allows a user to generate SEM descriptions for a video and upload the video and the SEM description to the CMSS. On the consumer side, SESim and SEMP can be used for simulating and rendering, respectively, the effects that are contained in the SEM description. UPnP A/V support (including the bundling feature) is currently implemented in SESim only, and not yet in SEMP. Therefore, SEMP is currently not able to browse or download files via UPnP A/V but it is able to render available SEM descriptions and videos.

The first user study that was presented in Section 3 shows that sensory effects increase the user experience. The bundling functionality of the CMSS is a way to propagate this information within the UPnP A/V network. The CMSS supports the consumption of MPEG-V SEM descriptions as follows: the SEM description is bundled with the video, as described in Section 2. Both, the Sensory Effect Metadata and the video are still stored separately in the CMSS but the user has only to browse one of the two files (e.g., only the video). Information about the bundled file is advertised in the DIDL-Lite description and can be used to download both files at the same time.

The second user study in Section 3 concludes that the perceived video quality can be increased by using sensory effects. Therefore, the storage requirements of the CMSS can be reduced by decreasing the bit rate of the video and instead supplying sensory effects. As a result, the network load and bandwidth requirements for retrieving the content from the CMSS are also reduced. Results showed that the video sequence *Earth* with a bit rate of 2204 kb/s and accompanied by a SEM description was perceived as being better than the same video sequence with a bit rate of

6701 kb/s without a SEM description. The used SEM description had a file size of approximately 3 Kbytes. Hence, the bit rate can be reduced by approximately 66% by using Sensory Effects, without decreasing the perceived video quality.

Note that there is a large variation in bit rates and perceived quality for other video sequences. Furthermore, for other video sequences the reduction can be much less than for this example but can still be feasible. For more detailed results refer to [17].

5 Conclusion and Future Work

In this paper we presented a metadata-based content management and sharing system which allows modifications, extensions and replacements in various parts of the system. Furthermore, the paper described a user-centric configuration that enables the user to create an individual configuration for displaying her/his content. Further, we introduced an extension of UPnP A/V for bundling content. This feature is an alternative to container formats and can be used to associate metadata or multimedia content which are stored in separate files. Moreover, the paper presented MPEG-V and how the CMSS can deliver bundled SEM descriptions within a UPnP A/V network in order to enrich the user experience. Finally, the CMSS is one of the first applications that support the enrichment of video content by providing bundled SEM descriptions via UPnP A/V.

Future work items are the integration of an adaptation and transcoding unit for providing the most suitable format (e.g., in terms of quality, resolution, bit rate) for various devices. Furthermore, UPnP A/V does not support user profiles which enable every user to have her/his own view on the content and preferences. A possible solution for this problem would be to integrate UPnP-UP (UPnP User Profile) as described in [20]. Another future work item would be the evaluation of the current system, for example, an evaluation on the question: "Is it faster and easier to find a specific content item via customized browsing than via searching?" Additionally, SEMP will be extended to support the retrieval of videos and SEM descriptions via UPnP A/V.

Acknowledgments. This work was supported in part by the EC in the context of the NoE INTERMEDIA (NoE 038419).

References

1. Gantz, J.F., et al.: The Expanding Digital Universe: A Forecast of Worldwide Information Growth Through 2010. IDC White Paper (2007)
2. UPnP Forum: MediaServer V 1.0 and MediaRenderer V 1.0, `http://upnp.org/standardizeddcps/mediaserver.asp`
3. Schulzrinne, H., Casner, S., Frederick, R., Jacobson, V.: RTP: A Transport Protocol for Real-Time Applications. RFC 3550. Internet Engineering Task Force (2003), `http://www.ietf.org/rfc/rfc3550.txt`

4. Schulzrinne, H., Rao, A., Lanphier, R.: Real Time Streaming Protocol (RTSP). RFC 2326. Internet Engineering Task Force (1998), http://www.ietf.org/rfc/rfc2326.txt
5. JBoss Application Server, http://www.jboss.org
6. MySQL, http://www.mysql.com/
7. VideoLan, http://www.videolan.org
8. CyberGarage: CyberLink for Java, http://www.cybergarage.org/cgi-bin/twiki/view/Main/CyberLinkForJava
9. Mime-util (Mime Type Detection Utility), http://sourceforge.net/projects/mime-util/
10. GNU Libextractor, http://gnunet.org/libextractor/
11. FFmpeg, http://ffmpeg.org/
12. TheMovieDB.org, http://www.themoviedb.org/
13. MusicBrainz, http://musicbrainz.org/
14. Debique, K., Igarashi, T., Kou, S., Moonen, J., Ritchie, J., Schults, G., Walker, M.: ContentDirectory:1 Service Template Version 1.01 (2002), http://www.upnp.org/standardizeddcps/documents/ContentDirectory1.0.pdf
15. ISO/IEC 23005-3 FCD: Information technology – Media context and control – Sensory information. ISO/IEC JTC 1/SC 29/WG 11/N10987, Xi'an, China (2009)
16. Waltl, M., Timmerer, C., Hellwagner, H.: Increasing the User Experience of Multimedia Presentations with Sensory Effects. In: 11th Int'l. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2010), Desenzano del Garda, Italy (2010)
17. Waltl, M., Timmerer, C., Hellwagner, H.: Improving the Quality of Multimedia Experience through Sensory Effects. In: 2nd Int'l. Workshop on Quality of Multimedia Experience (QoMEX 2010), Trondheim, Norway (2010)
18. Waltl, M., Timmerer, C., Hellwagner, H.: A Test-Bed for Quality of Multimedia Experience Evaluation of Sensory Effects. In: 1st Int'l. Workshop on Quality of Multimedia Experience (QoMEX 2009), San Diego, USA (2009)
19. amBX UK Ltd., http://www.ambx.com
20. Sales, T.B.M.d., Sales, L.M.d., Pereira, M., Almeida, H., Perkusich, A., Gorgônio, K., Sales Jr., M.A.d.: Towards the UPnP-UP: Enabling User Profile to Support Customized Services in UPnP Networks. In: 2nd Int'l. Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM 2008), Washington DC, USA (2008)