ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Eiji Kasutani

New Frontiers in Universal Multimedia Access

ITS Report 04.22

Lausanne, EPFL
September 2004

Abstract

Universal Multimedia Access (UMA) refers to the ability to access by any user to the desired multimedia content(s) over any type of network with any device from anywhere and anytime. UMA is a key framework for multimedia content delivery service using metadata.

This report consists of three parts. The first part of this report analyzes the state-of-the-art technologies in UMA, identifies the key issues and gives what are the new challenges that still remain to be resolved in UMA. The key issues in UMA include the adaptation of multimedia contents to bridge the gap between content creation and consuming, standardized metadata description that facilitates the adaptation (e.g. MPEG-7, MPEG-21 DIA, CC/PP), and UMA system designing considering its target application.

The second part introduces our approach towards these challenges; how to jointly adapt multimedia contents including different modalities and balance their presentation in an optimal way. A scheme for adapting audiovisual contents and its metadata (text) to any screen is proposed to provide the best experience in browsing the desired content. The adaptation process is modeled as an optimization problem of the total value of the content provided to the user. The total content value is optimized by jointly controlling the balance between video and metadata presentation, the transformation of the video content, and the amount of the metadata to be presented. Experimental results show that the proposed adaptation scheme enables users to browse audiovisual contents with their metadata optimized to the screen size of their devices.

The last part reports some potential UMA applications especially focusing on a universal access application to TV news archives as an example.

Contents

List of figures

List of tables

Part I: Universal Multimedia Access
- what are the challenges? -

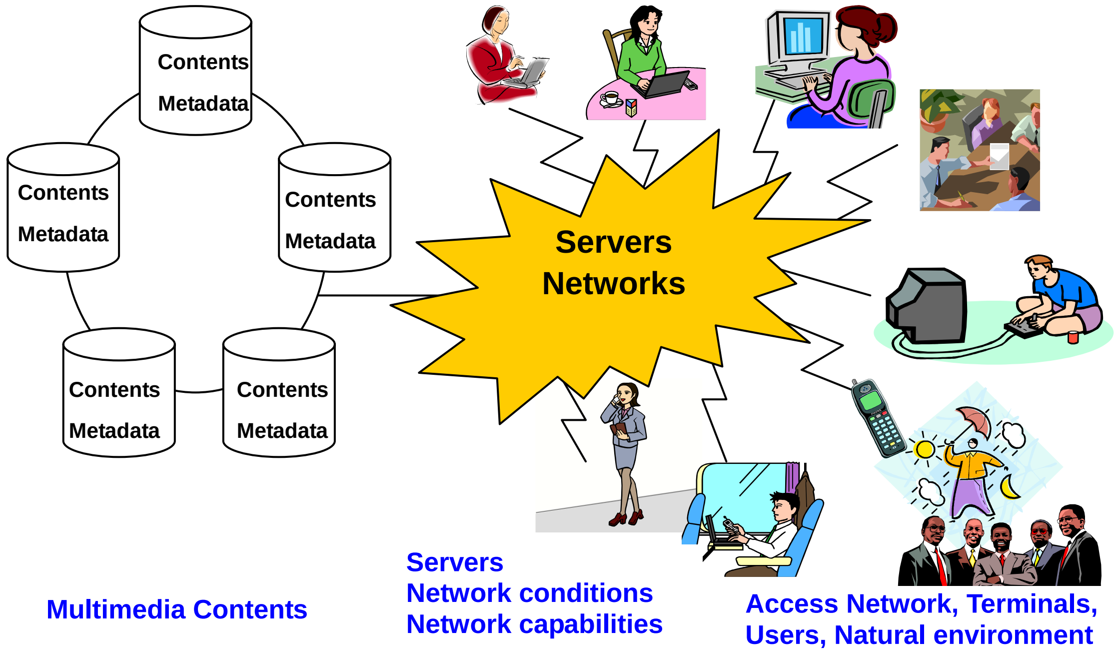


Figure 1. Multimedia content access through different environments.

1. Introduction

The recent advancement of multimedia technology has made content providers and consumers available numerous opportunities of coding, access and distribution. The efforts on standardization of multimedia content coding like MPEG-1/2/4, H.261, H.263, and H.264 enabled easy creation and distribution of contents. Thanks to this coding standard and the advancement of computational power, the number of digital appliances to consume multimedia contents, especially PC, DVD recordable player, PDA and mobile phones, is now increasing tremendously.

At the same time, the growth of the communication infrastructure has enabled access to information and multimedia services from almost anywhere at anytime. Various access networks (e.g. Ethernet, Bluetooth, wireless connections, 3G mobile, ISDN, xDSL, GPRS) and servers make it possible the access to a single or even distributed multimedia contents.

These advancements made a huge amount of multimedia contents accessible through various networks. Figure 1 illustrates the current circumstance. Under this circumstance, it is essential to allow access to the desired multimedia contents by different users with different terminals under various environments via various networks and servers. However, the main problem is that unless many individual technologies for multimedia consumption and network access are already present, there is still no solution that allows access of all types of data for all types of users in all types of conditions. Thus, interoperable solutions that enable access to services of different communities absorbing their differences have become an urgent and hot topic.

Universal Multimedia Access (UMA) refers to the ability to access by any user to any multimedia content over any type of network with any device from anywhere and anytime. This part analyzes the state-of-the-art technologies in UMA, identifies the key issues of UMA and new challenges that still remain to be resolved in UMA.

2. Universal Multimedia Access (UMA)

This section gives the overview of Universal Multimedia Access. The concept of Universal Multimedia Access (UMA) is explained in 2.1 to make clear the problem. Section 2.2 identifies who are the actors and which are the necessary tools in UMA systems. Key issues in UMA are given in 2.3.

2.1. Universal Multimedia Access concept

The concept of UMA is to enable access to any multimedia content over any type of network with any device from anywhere and anytime (universally). The initial motivation of UMA was to enable terminals with limited communication processing, storage and display capabilities to access rich multimedia contents. Some definitions of UMA are described in the following paragraph.

- The concept of UMA is to enable access to any multimedia content over any type of network, such as Internet, Wireless LAN or others, from any type of terminals with varying capabilities such as mobile phones, personal computers, and television sets [Mohan99].

- Universal multimedia access (UMA) deals with delivery of multimedia contents under different network conditions, user and publisher preferences, and capabilities of terminal devices [Perkis01].

- UMA refers to the framework where information is accessed in a suitable form and modality under the current complex and dynamic usage environment such as devices, networks, terminals, preferences, personalization, and other factors of usage environment [Vetro03].

- The primary function of UMA services is to provide the best QoS or User experience by either selecting appropriate content formats, or adapting the content format directly, to meet the playback environment, or to adapt the content playback environment to accommodate the content [SumISCAS03].

- Universal Multimedia Access (UMA): The notion (and associated technologies enabling) that any content should be available anytime, anywhere, even if after adaptation. This may require that content be transcoded form, for example, one bit rate or format to another or transcoded across modalities; e.g., text to speech. UMA concentrates on altering the content to meet the limitations of a user's terminal or network [Pereira03].

- Universal Multimedia Experience (UME): The notion that a user should have an equivalent, informative experience anytime, anywhere. Typically, such an experience will consist of multiple forms of multimedia content. Each will be adapted as in UMA but rather than to the limits of equipment, to limits that ensure the user has a worthwhile, informative experience. Thus, the user is central and the terminal and network are purely vehicles of the constituent content [Pereira03].

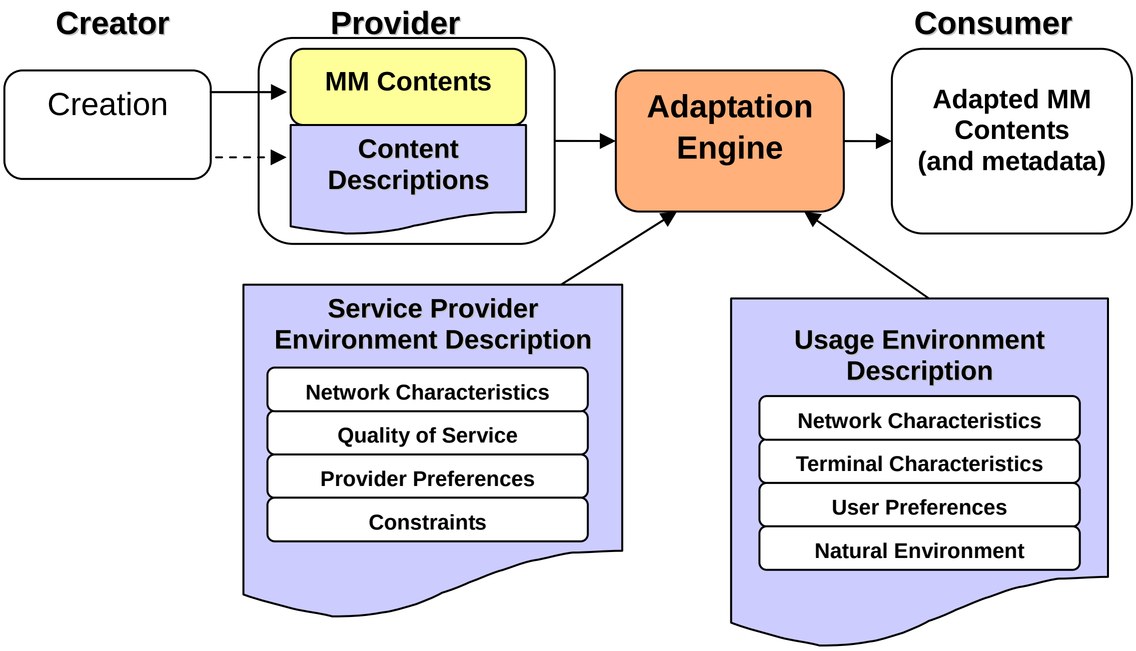


Figure 2. Overview of Universal Multimedia Access System.

Today, the UMA scope includes not only adaptation to terminals or networks but also includes **adaptation to all actors in UMA services, which includes the creator, provider and the consumer of the content, to maximize their satisfaction**. The most relevant emerging trend in UMA is **User-centric multimedia content adaptation**, instead of terminal centric adaptation.

2.2. Overview of UMA system

Figure 2 illustrates the overview of a UMA enabled system (we call 'UMA system'). A UMA system is a system which enables the consumer to access to the desired contents.

The main actors in UMA systems are as follows:

- **Creator**: The people or organization who creates the multimedia contents.
- **Provider**: The people or organization who provides multimedia content delivery service.
- **Consumer**: The people or organization who consumes the multimedia contents.

For example, the Creator creates a movie film, the Provider delivers this film and the Consumer watches it. Professional broadcast programs could involve several Creators and Providers, while personal home video could be created and delivered by the same person. As already mentioned in 2.1, the main objective in UMA systems is to maximize satisfaction of all these actors.

To achieve UMA systems, multimedia contents must be adapted to the Consumer especially in terms of terminal and network conditions and user characteristics. For easier adaptation of the desired content to the Consumer, it is preferable to have descriptions to fill the gap between media format and terminal, network, user characteristics. Descriptions required for UMA systems are described in the following:

- **Content Description**: information of the contents. Title, rights, structure, media format, variation, resolution, color, language, features, etc. (e.g. MPEG-7 tools)

- **Service Provider environment description**: quality of service (e.g. streaming bandwidth), constraints, privacy policy, network characteristics, limitations on distribution, etc.

- **Usage environment description**: descriptive information about various dimensions of the usage environment of the consumer to accommodate, for example, the adaptation of multimedia contents for transmission, storage and consumption like access network characteristics (eg. available bandwidth, packet loss rate), Terminal characteristics(eg. screen size, CPU power, available decoders), User Preference, Natural Environment (eg. location, time, weather, temperature) .

The Universal Multimedia Access concept involves the idea of content adaptation based on those descriptions of the content, service provider and the user environment. An **Adaptation Engine** is essential to bridge the gap between content creation and consuming. The content is adapted to the Consumer considering all these three types of descriptions. This adaptation engine would have a functionality to select the best variation for the consumer, to transform the media format to adapt the network conditions and the device capability of the consumer, and to deliver the preferable content in preferable mode for the consumer.

From Figure 1 and Figure 2, we can conclude that the key of UMA mainly exists in the **adaptation between provider and consumer** in order to maximize the quality of service and experience for both of them. For adaptation of the content to the consumer, three types of descriptions are necessary, which are multimedia content description, service provider environment description and a user environment description.

At the same time, it is essential to decide how to locate all these components, the adaptation engines, contents and descriptions, in **UMA system designing** considering the available computational resources. Besides the computational resources, the protection of both the **consumers' privacy** and the **provider and content holders' rights** of the contents are very important in UMA systems to make the system practical and to maximize the quality of service and experience for all actors. Content description, adaptation and privacy/rights management relies quite a lot on each other and also on the target application. Thus, it is quite important to consider all of these aspects in the beginning of a UMA system designing process instead of developing each of them separately and combining them all afterwards.

2.3. Key issues in UMA

This subsection indicates the key issues in UMA. As described in 2.2, the key issues in UMA can be concluded as follows;

A) Adaptation engine.
B) (Standardized) Description for adaptation.
C) UMA system designing.
   - Privacy protection of consumer (and provider).
   - Rights management of service provider and content holder.
   - UMA application

2.3.1. Adaptation engine (Adaptation between provider and consumer)

The major problem in UMA systems is the adaptation between provider and consumer that maximizes the quality of service and user experience. From the consumer side, UMA allows consumers access to a rich set of multimedia content through various connections such as Internet, Ethernet, DSL, Wireless LAN, Cable, Satellite, broadcasting and others, with different terminal devices. From the content or service provider side, UMA promises to deliver timely multimedia contents with various formats to wide range receivers that have different capabilities and are connected through various access networks [SumISCAS03]. In both cases, the adaptation engine should bridge the gap between media format and terminal, network, user and provider characteristics. This adaptation engine could contain selection of the best variation for the consumer, transformation of the media format to adapt the network conditions and the device capability of the consumer, and deliver of the preferable content for the consumer. Details are given in section 3.

2.3.2. (Standardized) Description for adaptation

For adaptation of the content to the consumer, three types of descriptions, multimedia content description, service provider environment description and a user environment description, are necessary. These descriptions are strongly desired to be described in some standard format to allow different user communities to interact in an interoperable way.

Content descriptions can be added manually, semi-automatically, and fully automatically depending on their aspects. Semantic and subjective features like ‘movie title’ and ‘author’ should be added manually. Low-level features like color, texture and shape of the image and video can be extracted automatically. Technologies like scene change detection or visual object extraction tries to automate the content structure description process. There are also many researches which try to fill the semantic gap between low level features and the semantics of the contents, for example, linking an image object name with a combination of low level features. Other features like ‘actors in the movie’ can be extracted semi-automatically by implementing object tracking and face recognition technology. MPEG-7 standard plays a key role in providing a description of a content [MPEG7MDS][MPEG7Visual]. MPEG-7 tools that can be used for adaptation are described in section 4.

Service provider environment descriptions and usage environment descriptions should include factors that may influence the access to multimedia contents. Besides terminal and network characteristics, location information, user preferences, privacy policies, content delivery policies are some other features that may affect the type of delivered multimedia content should be included. These descriptions make it easier the adaptation to bridge the mismatch between provider and consumer. The emerging standard MPEG-21, especially Part 7 Digital Item Adaptation (DIA), aims at fixing these gaps by providing the standardized descriptions and tools that can be used by the adaptation engines [MPEG21][MPEG21DIA]. Details of MPEG-21 DIA are described in section 4.

2.3.3. UMA system designing

In UMA system designing, it is necessary to decide how to locate the adaptation engines, contents and descriptions considering the available computational resources. In addition to all of the aspects described in A) and B), the following three aspects should also be considered for UMA systems. It is quite important to consider all of these aspects in the beginning of a UMA system designing process instead of developing each of them separately and combining them all afterwards.

Privacy protection of Consumer:
Under the situation that a large amount of personal information (e.g. user preference, usage history, access information, location information, user's terminal) is required for content adaptation, privacy would become a big concern in UMA systems. Potential problems on privacy in UMA systems and activities relevant to privacy are analyzed in section 5.

Rights protection of Provider and/or content holder:
For UMA services it is essential to protect the value of the content and the right of the rights holders. On the other hand, interoperability is significant to realize an open multimedia infrastructure. MPEG-21 tries to give solutions to this problem [MPEG21IPMP]00.

UMA Application
The system designing of UMA systems quite depends on the application. In case where the computational resources of the client would be very limited, the adaptation engine has to be located on the provider side. On the other hand, if more flexibility in handling the contents by the user is necessary, then the adaptation engine can be located on the consumer device to avoid network request each time the user wants to deal with the contents. Challenges for designing the applications that allows the users to access, store and process information are emerging. Some UMA application and their system design are explained in section 6.

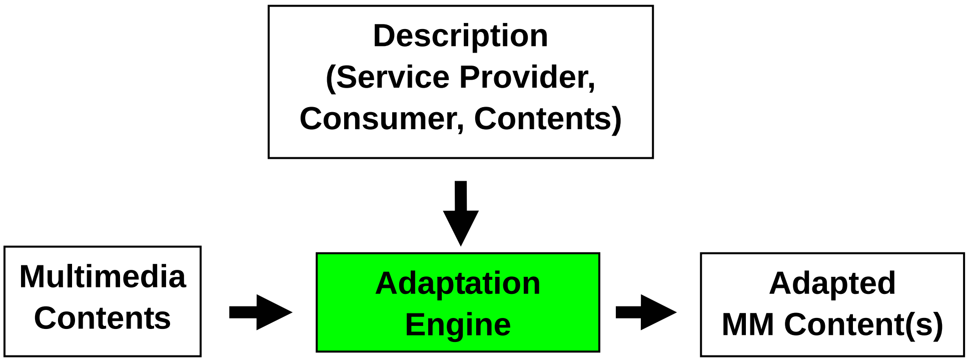


Figure 3. Content adaptation framework in UMA systems.

3. Content Adaptation in Universal Multimedia Access

This section provides a detailed analysis on how to perform content adaptation. It should be noted that the adaptation is achieved based on descriptions mentioned in section 2, which are content description, service provider environment description and usage environment description. Those descriptions are quite useful to enable efficient and appropriate adaptation of the contents. With all descriptions available, the adaptation engine adapts the content by transforming it on the fly, by selecting the content variation or by selecting the preferred content in a way that the best possible experience is provided to the consumer.

3.1. Content adaptation framework in UMA systems

Various contents in various formats are delivered by various service providers via various servers and networks to various terminals of various users. Therefore, a large number of parameters of the contents, service providers and usage environments need to be taken into consideration for adaptation. Some examples of adaptation parameters are shown in Table 1.

Table 1. Examples of parameters for adaptation.

Content	Service Provider	Usage Environment
Media formats, Bitrate	Quality of Service	Access Network (Bandwidth)
Spatial resolution	Available Bandwidth	Display resolution / color
Temporal resolution	Error rate	Memory / CPU / Decoders
Number of Colors	Constraints	User preference
Limitations, rights	Delay	Access location, time

Figure 3 illustrates a diagram of content adaptation framework in UMA systems. The multimedia contents and all the descriptions are input into the adaptation engine, and the adaptation engine adapts the contents by selection and transformation based on all descriptions to form a multimedia content adapted to the consumer.

Table 2. Relation between data to access and adaptation engines.

	*Source	*Variations	Required engines
Type 1	1	1	**Transformation engine**
Type 2	1	m (>1)	**Variation selection engine** (+ Transformation engine)
Type 3	n (>1)	1	**Content selection engine** (+ Transformation engine)
Type 4	n (>1)	m (>1)	Content selection engine + variation selection engine (+ Transformation engine)

*Source: Original source content.

*Variation: Set of contents derived from a single source content.

3.2. Adaptation engines in UMA systems

In this subsection we analyze which types of adaptation engines are required in UMA systems. There are numerous ways to categorize adaptation type, for example, if the adaptation should be processed in real-time or not, if the target application is push or pull, if the adaptation are performed automatically or manually, if the adaptation process requires a lot of computation or not, if the adaptation is in signal level, perception level or semantic level, and so on.

We focus on what type of multimedia contents the consumer accesses. The first point is if the data to access consists of just one source or various sources. The second point is if there are variations of each content created from a single source or not. Several alternative versions or variations derived from a single source may exist (e.g. same program with different frame rate, resolution, number of colors, languages, etc).

Table 2 shows the relation between data to access and required adaptation engines. We divide into four types of data to access and describe the adaptation engines required for each of them. The first type, where there is only a single content with no variation, requires a transformation engine that transforms the target content into an adapted content. The second type, where there is a single content with multiple variations, requires a variation selection engine that selects the best variation among all variations. After selecting the best variation, it is also possible to have a transformation engine for better adaptation. The third type, where there are multiple sources with no variation, requires a content selection engine that selects the best content(s) among all programs. After selecting the best program, it is also possible to have a transformation engine for better adaptation. The fourth type, where there are multiple sources with multiple variations, requires both a content selection engine and a variation selection engine that selects the best variation of the best content. After selecting the best content and variation, it is also possible to have a transformation engine for better adaptation. The details of those engines are explained in the following subsections, from 3.2.1 to 3.2.4.

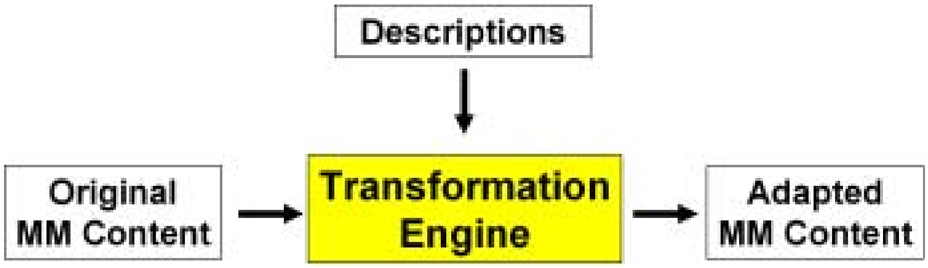


Figure 4. Content adaptation by transformation.

3.2.1. Transformation engine

Figure 4 illustrates the diagram of the content adaptation process by transformation engine. The transformation engine transforms a single multimedia content into an adapted multimedia content on-the-fly. The descriptions support the adaptation process to provide the best experience to the user and to reduce calculation cost enough to achieve a real-time adaptation. One relevant application could be broadcast video content delivery to mobile users [Bjork00][Vetro01].

The advantages of content adaptation by transformation are that the storage cost is low because only a single content variation needs to be stored and that an accurate adaptation to every device can be performed considering the device capability in real-time. On the other hand, the drawbacks are that the engine could have only limited operations that enable real-time or low delay processing, unless it would cause long delay or an engine with quite a lot of computational power would be necessary.

Some content transformation examples are described in the following:

- Adjustment of network capabilities (e.g. bandwidth, delay, error rate) and terminal capabilities (e.g. screen size, terminal power, memory, CPU, decoder) between the service provider and consumer.
- Adjustment of content presentation (visualization) based on user preferences (e.g. preferred mode(e.g. small image high frame-rate or large image low frame-rate, all content or summarized content), difficulties in vision or hearing) and natural environments (e.g. location, time, weather, color temperature adjustment)
- Content (and metadata) visualization for browsing (e.g. key-frame visualization)
- Content presentation based on service provider environment (e.g. limitation on playing mode and time).
- Content summarization (e.g. 1 hour news in 5 minutes).
- Content digest, highlight (e.g. goal scene in a football match).

What to transform of the content includes spatial resolution, temporal resolution, modality, content length, coding format, coding parameters, presentation, content length, spatial region, temporal region, color and properties. The contents are transformed in the way that maximizes the experience of the consumer.

Related works for transformation engines are shown in the following:

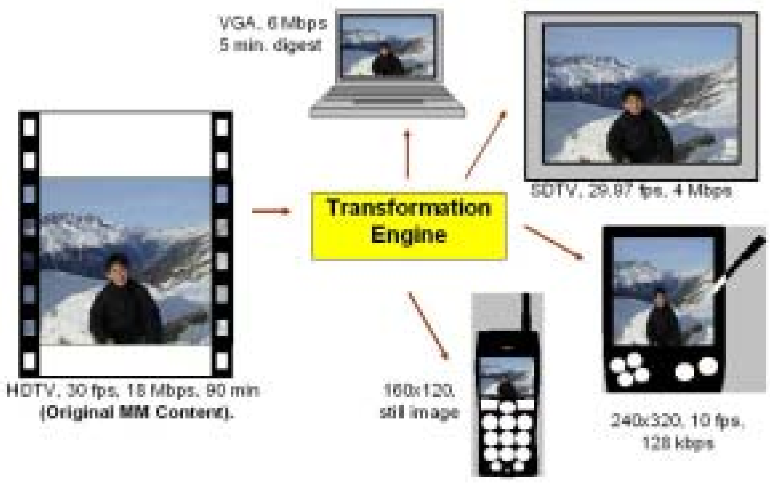


Figure 5. Example of content adaptation by transformation

Conversion of spatio-temporal resolution

- Convert the spatial resolution and color depth of the image/video
  - ✧ e.g. VGA to QVGA, NTSC to PAL, 24bit to 8bit color depth, color to grayscale, etc..
- Convert the temporal resolution of the video
  - ✧ e.g. 30fps to 24 bps(NTSC to PAL), 30 fps to 10 fps, etc.

Bitstream transcoding

- Conversion of video coding formats
  - ✧ e.g. MPEG-2 to MPEG-4, MPEG-4 to WMF, …
- Coding parameter conversion [SumISCAS03]
  - ✧ bit rate (e.g. 6.0 Mbps of TV broadcast to 128 kbps for mobile phones)
  - ✧ frame-rate (e.g. 30 fps to 24 fps)
  - ✧ spatial resolution (e.g. CIF to QCIF)
  - ✧ DCT coefficients, quantization level, error resilience, etc..

Transmoding

- Transformation of modalities.
  - ✧ video-to-image conversion (mosaicing, key frame extraction)
  - ✧ image-to-text conversion etc.) [Cavallaro03].
  - ✧ Image-to-video conversion [Bruijin02]
  - ✧ speech-to-text conversion
  - ✧ SVG-to-image conversion [Lin04], etc..

Extraction of spatio-temporal segments

- Primary object extraction,
- important video segment extraction,
- Use of Region of Interest (ROI). [Lee01], etc..

Summarization

- Control of playback speed (e.g. 1.5 times fast play,)
- Semantic level summarization (pick-up high light scenes in sports video)
- Summarize 1 hour news in 5 minutes, etc.

Visualization

- e.g. preferred news category visualization and selection, key-frame browsing for understanding the video contents, etc.

Personalization

- All of the above transformation based on user's interest.
  - transformation to preferred resolution, format, modality, regions/segments…

Challenges for transformation engines are descried in the following:

- What kind of view does the user really want? (not only in terms of terminal and network)
- How should the contents be presented to the user?
- How to allow the users easy access to the contents?
- How to consider the tradeoff between complexity and quality?
- What kinds of description are necessary for real-time transformation?
- How to evaluate the quality of adaptation and experience? (Quality metrics)
- Quality of service measurements.
- Quality measure of the value of the received contents for the user.
- (Terminal) Capability negotiation.

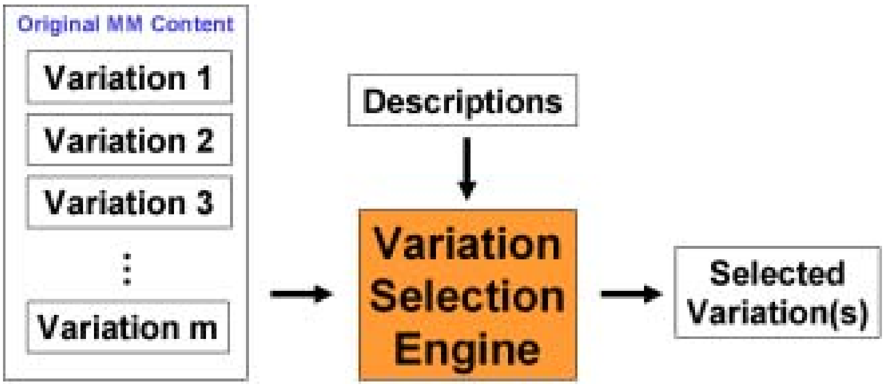


Figure 6. Content adaptation by variation selection.

3.2.2. Variation selection engine

Figure 6 illustrates the diagram of the content adaptation process by variation selection engine. The variation selection engine selects the best variation from all variations delivered from a single multimedia content. The descriptions support the selection process to provide the best experience to the user. The most typical example of content adaptation by variation selection would be a web site preparing several variations for each type of terminal (PC, PDA, mobile phone) and delivering automatically the most appropriate one by analyzing environmental variables (e.g. HTTP_USER_AGENT).

The advantages of content adaptation by selection are that the adaptation process is very quick because the system just has to select the best variation and deliver it to the consumer. At the same time, the system needs just a small amount of computational resources for adaptation. On the other hand, the main drawback is that the engine could have only limited variations and sometimes there would be no content that fits the user condition. It is possible to solve this problem by applying a transformation engine to the selected variation in order to increase the consumer's experience. Other problems are that the storage cost and creation cost becomes higher. Furthermore, the management cost is high. If there are some changes on the original content, all of the variations also need to be changed, which requires quite a lot of time and power.

Variation selection examples are basically the same as those of transformation engine. The main difference is that the variations are created beforehand. What to select includes best spatial resolution, temporal resolution, modality, content length, coding format, coding parameters, presentation, content length, important spatial region, temporal region, color selection for the user, and so on. Some content variation selection examples are described in the following:

Variation selection based on terminal and network capabilities.

- Selection of the best variation by adjusting network capabilities (e.g. bandwidth, delay, error rate) and terminal capabilities (e.g. screen size, terminal power, memory, CPU, decoder) between the service provider and consumer.

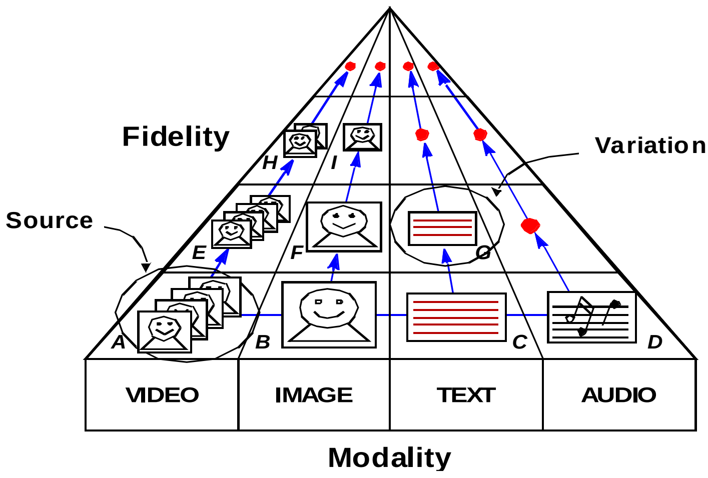


Figure 7. Illustration of different variations of a single source content.

Variation selection based on user preferences and natural environments.

- Selection of the best variation based on user preferences (e.g. preferred mode(e.g. image quality or smoothness), desired content length (all content or summarized content), difficulties in vision or hearing) and natural environments (e.g. location, time, weather).

Variation selection based on provider preferences and usage restrictions

- Selection of the best variation based on service provider preferences and content descriptions
  - ✧ content designed for only TV or mobile phone
  - ✧ limitations on access networks
  - ✧ variation allowed to be distributed to limited types of devices (no distribution to mobile phones. etc), etc.

Scalable media

- Spatial and temporal scalable coding (e.g. MPEG-4 FGS [Chung03][ChenICME02][Shaar02][ChenICCE02], JPEG2000). Scalable media includes several layers of different resolutions in the content itself. The best layer is selected considering the usage environments.

The most relevant work in this field would be InfoPyramid framework developed by IBM [Li98]. This framework enables to describe the associations or relationships between different variations of multimedia content. This supports content management by tracking the variations of multimedia content that result from various types of multimedia processing such as summarization, translation, reduction, revision, transcoding and so forth. This also supports Universal Multimedia Access by allowing the selection of the most appropriate variation of the multimedia content for the specific capabilities of the terminal devices, network conditions or user preferences.

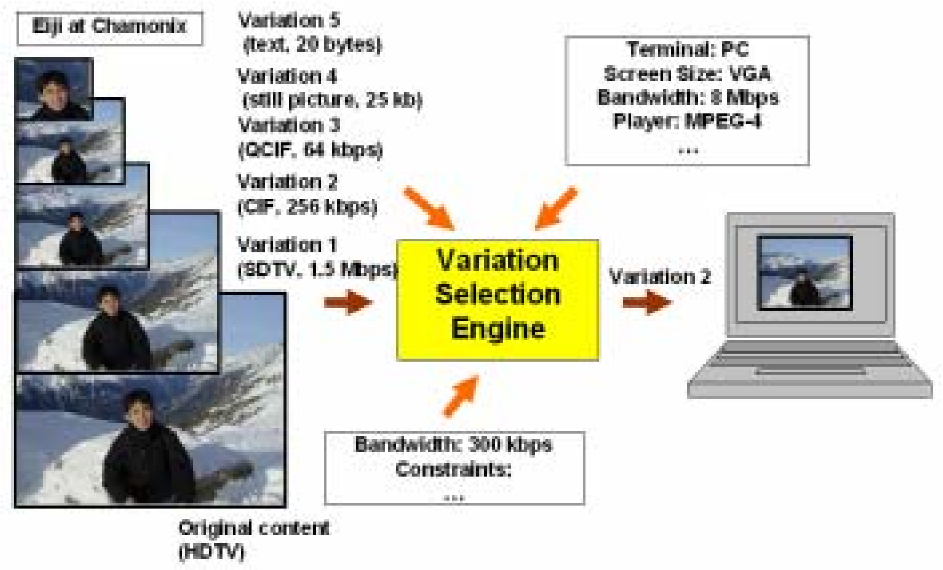


Figure 8. Example of content adaptation by variation selection.

Figure 7 illustrates a set of variations of multimedia content. The example shows the source video content in the lower left corner and shows eight variations: two variations are video content, three variations are images, two variations are text, and one variation is audio. Each variation has a fidelity value that indicates how close or faithful the variation content is to the source content [MPEG7MDS].

Currently, some web sites with visual contents already have variations in their servers, and allows user to select the contents considering their available bandwidth (eg. 56k or 300k?), and playable media format (e.g. "Real Player" or "Windows Media Player"?). As the numbers of different terminals are increasing, an automatic or semi-automatic selection method is emerging. Figure 8 illustrates an example of content adaptation by variation selection. In this example, the variation selection engine selects the best variation for the PC user considering his/her usage environment description, available content variations and restrictions of the service provider.

Challenges:

- What kinds of variations are necessary?
- How to generate variations?
- How to select the best modality?
- How to manage variations from a single content?
- How to evaluate the quality of selection? (Quality metrics)

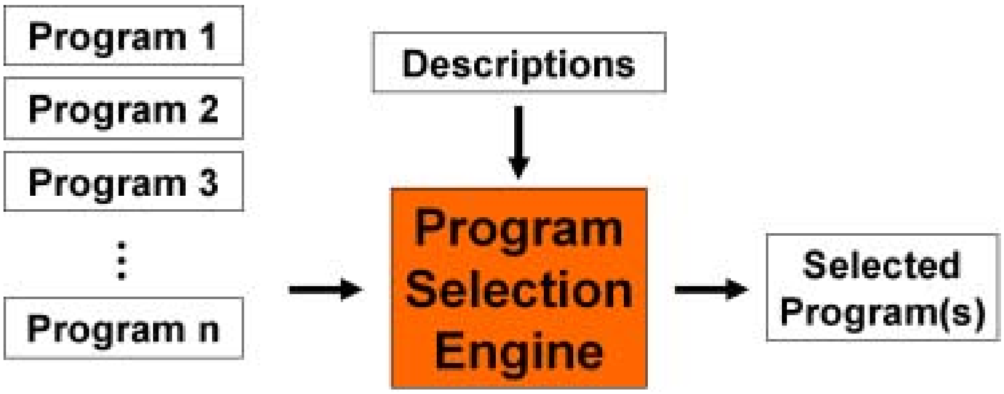


Figure 9. Content adaptation by program selection.

3.2.3. Content selection engine

Figure 9 illustrates the diagram of the content adaptation process by content selection engine. The content selection engine selects the best source content(s) for the consumer from all source contents (multimedia contents). The most typical example would be personalized content selection of TV programs, considering the user preference.

Functionality of content filtering

Some content selection examples are described in the following:

- **Recommendation** of TV programs considering user's age, location, sex, etc.
- **Selection of programs** that one would be interested
  - (e.g. which sports he/she likes, which type of movies he/she views frequently…).
- **Automatic recording** of interesting programs for a user
- **Gathering programs** of specific subject or topic
  - (e.g. collection of Swiss-related programs, browsing of some special news, collection of programs with some specific actors)
- **Restriction** of violent programs to children.

Content filtering techniques

Related works for program selection engines are listed up in the following:

- **Content filtering** [Angelides03]
  - rule-based filtering
  - content-based filtering
  - collaborative filtering.
- **User profiling** [Ferman03]
- **Usage history updating** [Ferman02],
- **Category profiling** (group of user profiling).

Angelides [Angelides03] divided content filtering technologies into the following three types. They can be used alone and also as a combination.

1) **Rule-based filtering** works with rules derived from statistics such as user demographics and initial user profiles. The rules determine the content that a user receives. Both the accuracy and the complexity of this filtering increase proportionally with the number of rules and the richness of the user profiles.

   Drawback:
   - It depends on users knowing in advance what content might interest them.
   - The accuracy and comprehensiveness of both the decision rules and the user modeling.

2) **Content-based filtering** chooses content with a high degree of similarity to the content requirements expressed either explicitly or implicitly by the user. Content recommendations rely heavily on previous recommendations. Hence, a user profile delimits a region of the content model from which all recommendations will be made.

   Drawback:
   - This filtering is simple and direct but it lacks serendipity;
   - Content that falls outside this region (and the user profile) could be relevant to a user but it won't be recommended.

3) **Collaborative filtering** are prediction algorithms over sparse data sets of user preferences. With collaborative filtering every user is assigned to a peer group whose members' content ratings in their user profiles correlate to the content ratings in the individual's user profile.

   Drawback:
   - Inclusion of new, unrated content in the model may take time before other users see and rate the content.
   - Also sometimes users who don't fit into any group end up being included because of unusual requirements.

The selected content(s) by the content selection engine can be transformed to adapt best to the consumer by using the transformation engine explained in 3.2.1.

There are another possibility to combine the content selection engine and transformation engine for better adaptation and selection. Typical examples would be collection of scenes with some actors, and browsing only goal scenes from all the weekend football matches. These examples cannot be achieved by only program selection, but scene-based evaluation and selection, transformation are required.

Challenges:
- Evaluation metrics for selection.
- Preparation and selection of descriptions for selection engine.
- Evaluation metrics of the selected results.
- How to update User profiles?

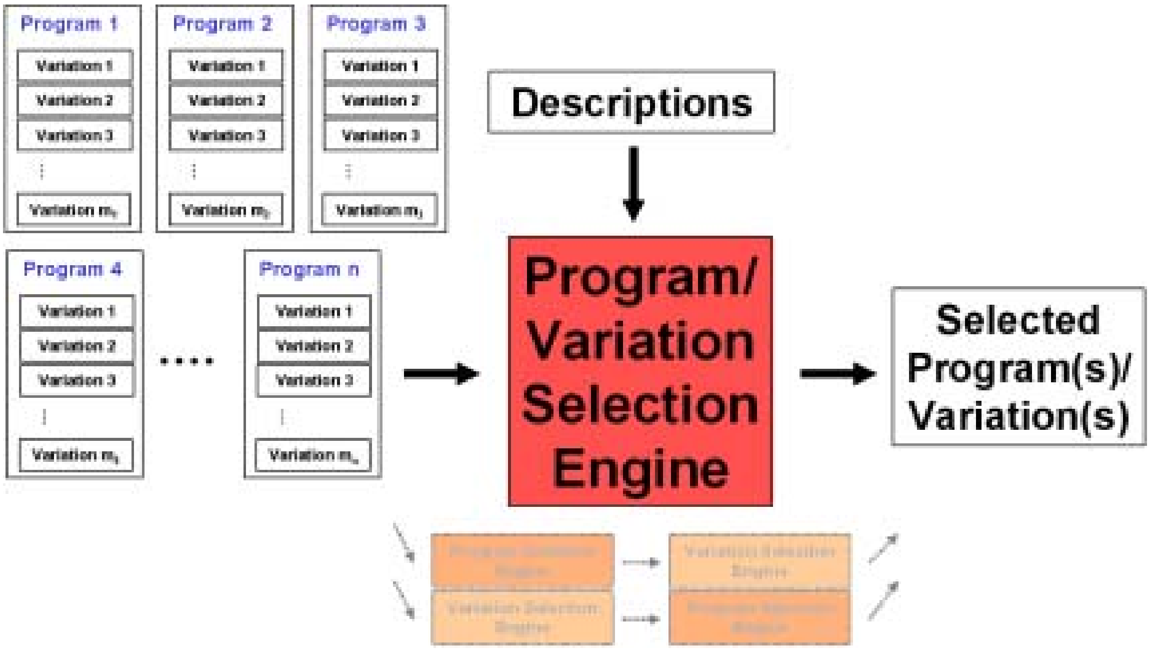


Figure 10. Content adaptation by program and variation selection.

3.2.4. Content and variation selection engine

Figure 10 illustrates the diagram of the content adaptation process by program and variation selection engine. The program and variation selection engine selects the best variation(s) from all programs and variations delivered from multiple multimedia contents. The descriptions support the selection process to provide the best experience to the user. The most typical example of content adaptation by program and variation selection would be listing and browsing available recorded programs from any device at anywhere anytime.

Program and variation selection examples are basically the same as those of program selection engine and variation selection engine. The main difference between program selection engines is that the variations are also created beforehand. The program and variation selection engine could include a program selection engine, variation selection engine and transformation engine.

Figure 11 illustrates a system with program and variation selection engine developed by Steiger [Steiger03]. A personalized multimedia content delivery system using user preferences and terminal/network capabilities are presented. Key issues of the system are content preparation (variation, MPEG-7 annotation tool), content adaptation and delivery using user/server preferences, terminal/network capabilities and usage history using MPEG-7 and MPEG-21 descriptions.

One of the main problems is that too many contents could be selected depending on the content database size or user's request or preference. Some visualization or presentation method of the selected contents is emerging to make it easier for the user to access the desired content. Figure 12 describes an example of a system with content/metadata visualization function.

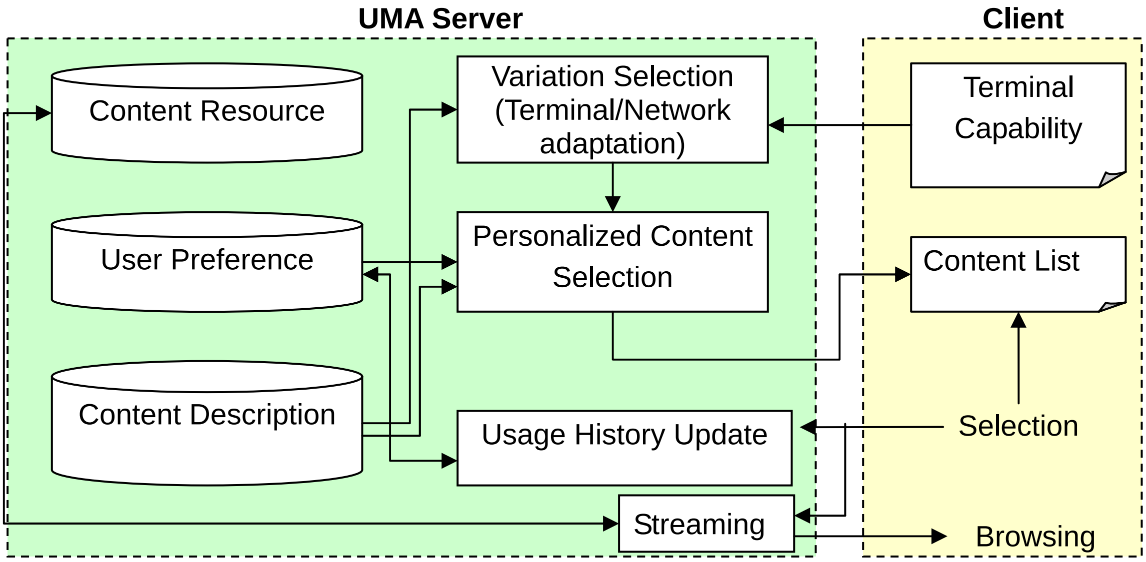


Figure 11. Personalized content delivery using program and variation selection engine.

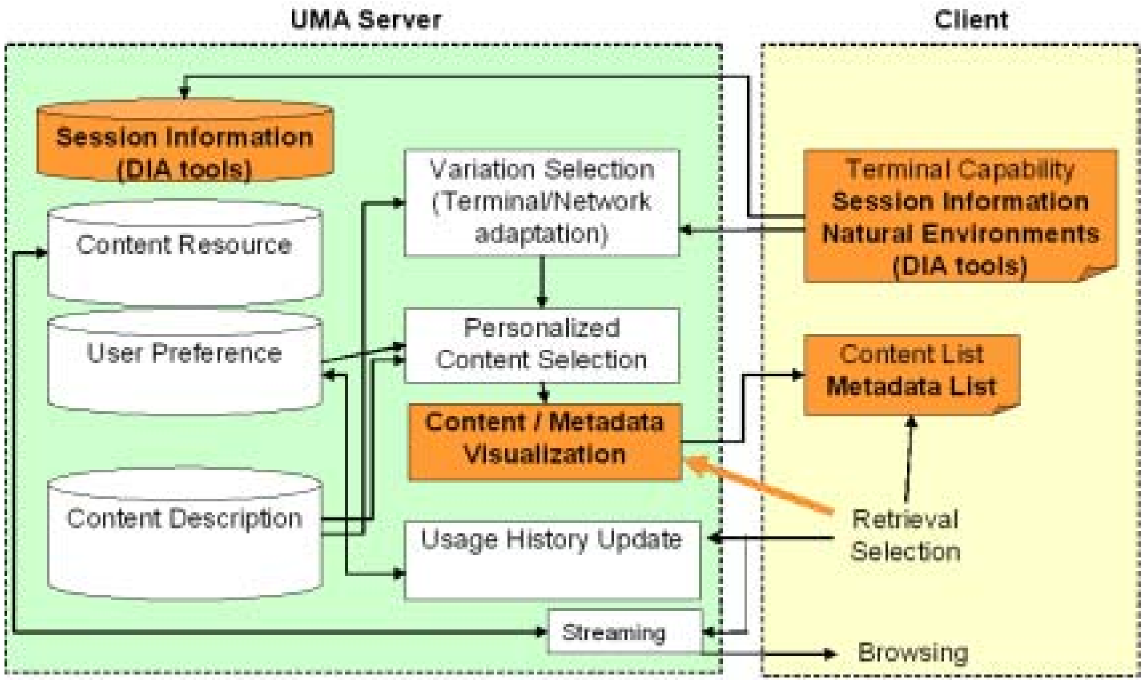


Figure 12. Example of Personalized content delivery with visualization functions.

3.3. Conclusion

This section presented a content adaptation framework in UMA systems and four types of engines that enable adaptation of the contents to the user. The adaptation engines adapt the content by transforming it on the fly, by selecting the content variation or by selecting the preferred content in a way that the best possible experience is provided to the consumer.

To achieve user-centric multimedia adaptation, there still remain a lot of problems. The problems can be divided into four categories; 1) Preparation of descriptions and variations, 2) Metrics for adaptation, 3) Presentation of the adapted contents, and 4) Evaluation metrics.

1) Preparation of descriptions and variations.

- What kinds of descriptions and variations are necessary for adaptation, for real-time transformation, for personalization, for transcoding, for a specific application?
- How to give an adequate keyword to describe the contents?
- How to manage variations from a single content?

MPEG-7 and MPEG-21 provide rich tools to describe contents, variations and environments. As the requirements for the system depends quite a lot on the application, it is essential to identify the necessary descriptions and variations for each application.

After identifying what kind of descriptions and variations are necessary, it is necessary to create them within a reasonable cost. It is also important how to update variations when there are changes in the source content.

2) Metrics for adaptation.

- How to consider the tradeoff between complexity and quality?
- How to select the best modality?
- How to update user profiles?
- Parameter configuration and evaluation metrics for personalization.
- How to create generic rules for adaptation?
- Evaluation metrics for transformation and selection.

3) Presentation of the adapted multimedia contents (and metadata).

The main problems are as follows;

- What kind of view does the user really want? (not only in terms of terminal and network)
- How should the contents be presented to the user?
- How to allow the users easy access to the contents?
- How to visualize the contents and/or their metadata for easy browsing?

Instead of adapting the content to the screen size of the consumer, the analysis on what kind of presentation the consumer really wants is necessary. Of course, the preferable way of presentation depends on the user.

After adaptation based on transformation or selection, there are cases that too many contents are selected and makes the user difficult to access to the desired content. Some assistance method for easy access like content visualization, metadata visualization is necessary. Video summary, content & metadata structuring, content visualization for increasing accessibility (shot/key frame presentation) are currently studied in many organizations, however, metadata visualization for increasing accessibility needs to be developed to increase accessibility (Figure 13)

4) Evaluation metrics.

- How to evaluate the quality of experience? (Quality metrics)
- Quality of service measurements.
- Quality measure of the value of the received contents for the user.
- How to evaluate the quality of selection and transformation?
- Evaluation metrics of the selected results.

The evaluation metrics of the obtained adapted contents for the user is emerging. PSNR or error rate, gained bit rate are not the appropriate way to measure the quality of experience provided by UMA services. The evaluation metrics should include some measurement of user satisfaction, which includes the quality of the context in the image/video, how much the provided contents fit their preferences, and any other factors that effects the user experiences.

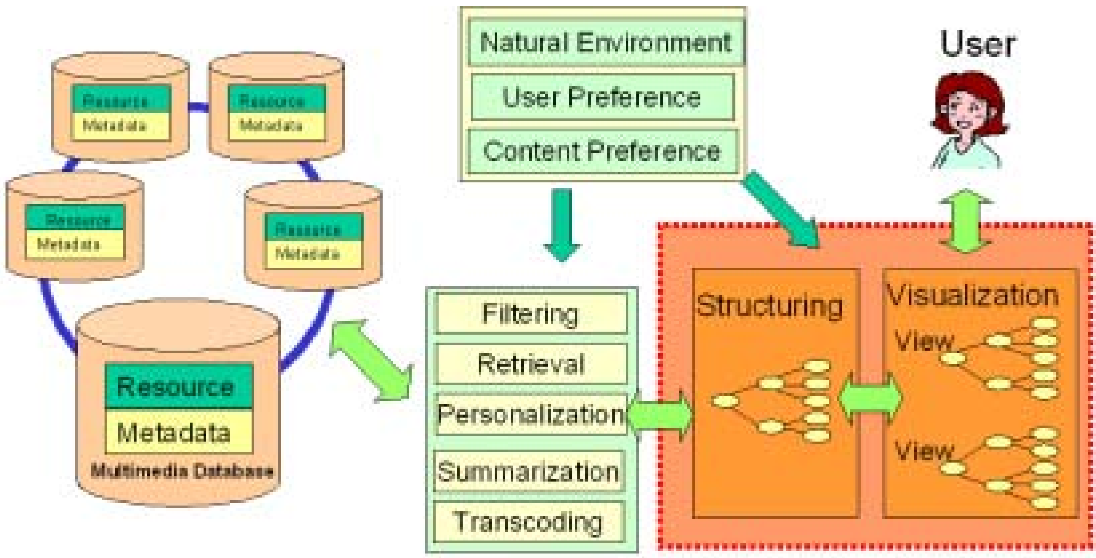


Figure 13. Metadata Visualization

4. Description for Adaptation

This section presents the state-of-the-art in descriptions for adaptation and tries to cover as many aspects that influence the multimedia content adaptation process as possible. As the interoperability among numerous contents, service providers and terminals are essential in UMA systems, these descriptions should be described in some standard format. MPEG-21, especially Part 7 Digital Item Adaptation (DIA) provides a rich set of standardized descriptions and tools necessary for adaptation. Some tools in MPEG-7 Part 5 Multimedia Description Scheme (MDS) are also important for Universal Multimedia Access. Besides MPEG, there are also standards relevant to adaptation. In this section, we explain in detail the MPEG-21 Part-7 DIA, MPEG-7 tools for adaptation and introduce other relevant standards to multimedia content adaptation to make clear what kind of aspects are important for adaptation.

4.1. MPEG-21 Part-7 Digital Item Adaptation

4.1.1. Overview of MPEG-21

The goal of MPEG-21 is to define the technology needed to support users to exchange, access, consume, trade, and otherwise manipulate Digital Items (DIs) in an efficient, transparent, and interoperable way [MPEG21]. MPEG-21 specifically takes into account digital rights management (DRM) requirements (called Intellectual Property Management and Protection (IPMP) in MPEG world) and targeting multimedia access and delivery using heterogeneous networks and terminals, which is quite related with Universal Multimedia Access.

"Digital Item" (DI) is a structured resources (such as video, audio, text, image, etc) with a standard representation, identification, and associated metadata within the MPEG-21 framework. Figure 14 shows the most important elements within this model, and how they are related. MPEG-21 Part-2 plays the role to define tools for declaring these Digital Items [MPEG21DID].

"User" is any entity that interacts within the MPEG-21 environment and/or makes use of DIs. Thus, a User of a system includes all members of the value chain (e.g., creator, rights holders, distributors (service providers) and consumers of Digital Items).

It should be noted that at this time (January 2004) the standardization of MPEG-21 is still ongoing. This report is based on DIS of MPEG-21 and there would be changes and modifications in the standard.

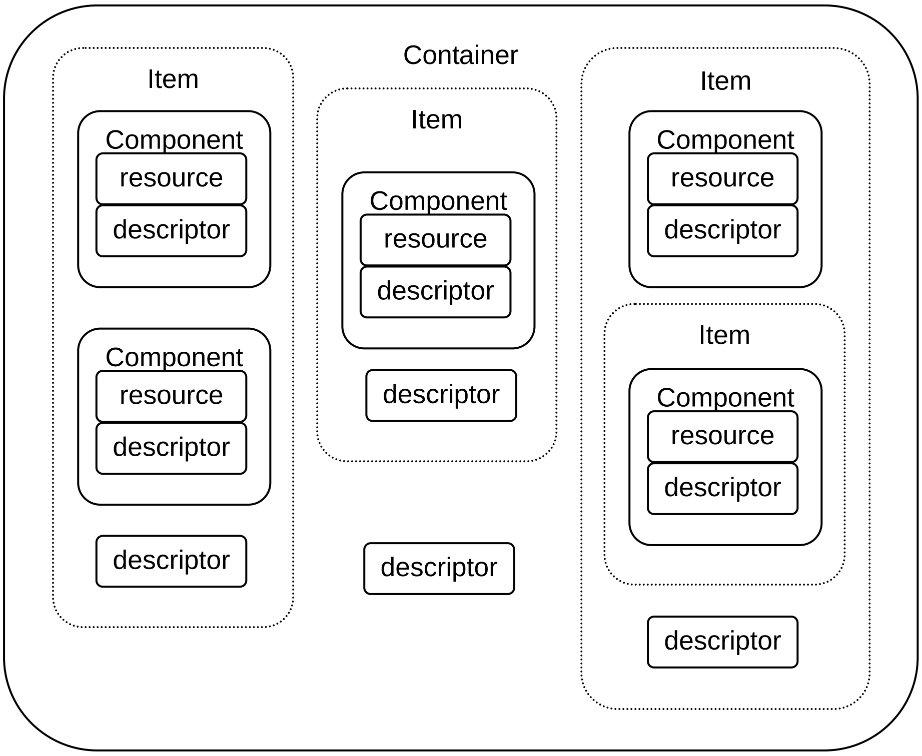


Figure 14. Relation between MPEG-21 elements

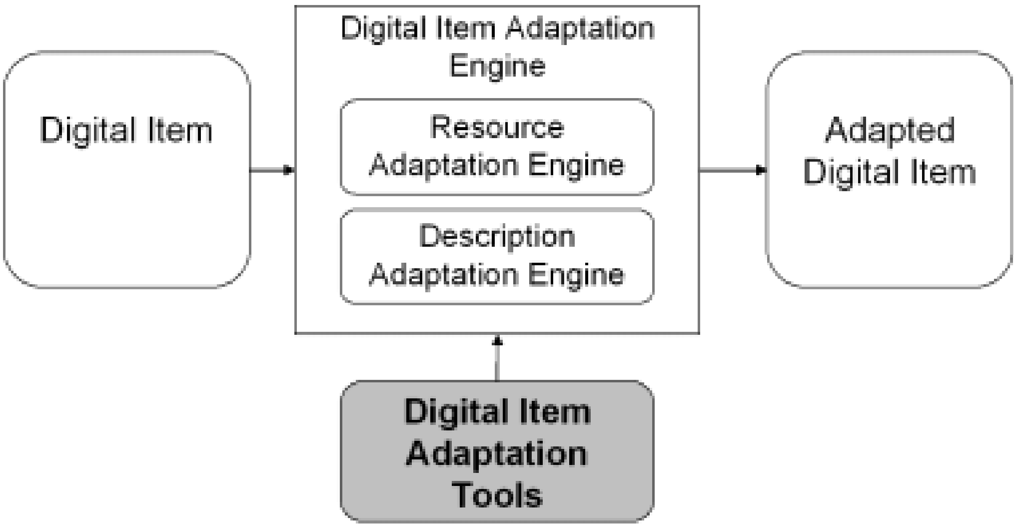


Figure 15. Illustration of MPEG-21 Digital Item Adaptation.

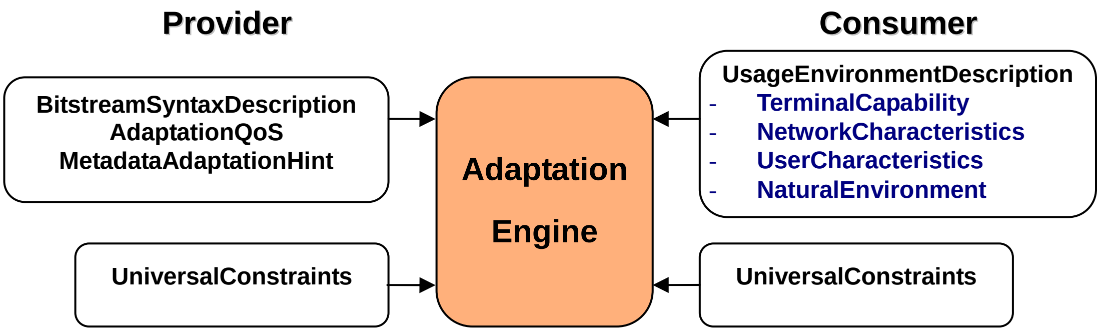


Figure 16. DIA tools in UMA systems.

4.1.2. Overview of MPEG-21 Part-7 Digital Item Adaptation

One goal of MPEG-21, especially Part-7 of MPEG-21, Digital Item Adaptation (DIA), is to provide standardized descriptions and tools that can be used by adaptation engines, which are quite relevant to Universal Multimedia Access [MPEG21DIA]. The conceptual architecture of MPEG-21 DIA is illustrated in Figure 15. DIs are subject to a resource adaptation engine, as well as a descriptor adaptation engine, which together produce the adapted DI.

Figure 16 shows the main DIA tools that are used as parameters for adaptation in UMA systems. The Digital Item Adaptation tools are divided into the following seven groups. The first five descriptions, which are quite relevant to UMA, are explained in detail in the following subsections, from 4.1.3 to 4.1.7.

1. Usage Environment Description Tools: Tools to describe various dimensions of the usage environment, which originate from Users to accommodate the adaptation of Digital Items for transmission, storage and consumption. It consists of Network, Terminal, User and Natural Environment description.

Descriptions tools from 2 to 4 are mainly for transcoding. These tools include a structure on the resource so it can be edited, and means for deciding on trading off parameters for QoS.

2. Bitstream Syntax Description: A BSD describes the syntax (high level structure) of a binary media resource. Using such a description, a Digital Item resource adaptation engine can transform the bitstream and the corresponding description using editing-style operations such as data truncation and simple modifications. It consists of two main technologies, Bistream Syntax Description Language (BSDL) and Generic Bitstream Syntax (gBS). BSDL is an XML schema based language to design specific bitstream syntax schemas for **particular media formats**. gBS schema is a generic schema enabling the construction of **resource format independent** bitstream syntax descriptions.

3. Terminal and Network Quality of Service: The AdaptationQoS description tools specified in this group describe the relationship between QoS constraints (e.g., on network bandwidth or a terminal's computational capabilities), feasible adaptation operations satisfying these constraints and associated media resource qualities that result from adaptation. The

AdaptationQoS descriptor therefore provides the means to trade-off these parameters with respect to quality so that an adaptation strategy can be formulated and optimal adaptation decisions can be made in constrained environments.

4. Universal Constraint Description tools: The Universal Constraints Description Tools enables the possibility to describe limitation and optimization constraints on adaptations.

Description 5 includes metadata adaptation hint information to reduce the complexity of adapting the metadata contained in a DI.

5. Metadata Adaptability: This description tool describes adaptation hint information pertaining to metadata within a digital item. This information is a set of syntactical elements with prior knowledge about the metadata that is useful for reducing the complexity of the metadata adaptation process. On the one hand they are used for filtering and scaling and on the other hand for integrating XML instances.

Descriptions 6 and 7 are descriptions of DIA tools to help the adaptation procedure. The former keeps the current state of interaction, and the latter describes the DIA descriptors required for a specific resource. These descriptions are distributed alone and are not associated with the resources.

6. Session Mobility: Session Mobility specifies tools to preserve a User’s current state of interaction with a Digital Item. The configuration state information that pertains to the consumption of a Digital Item on one device is transferred to a second device. This enables the Digital Item to be consumed on the second device in an adapted way.

7. DIA configuration: DIA Configuration provides a functionality to identify the DIA descriptors that are required for a specific resource, and to identify how choice/selections should be processed, e.g., displayed to Users or configured in the according to DIA descriptors, and identifies the location of the adaptation, e.g., receiver side, sender side or either side.

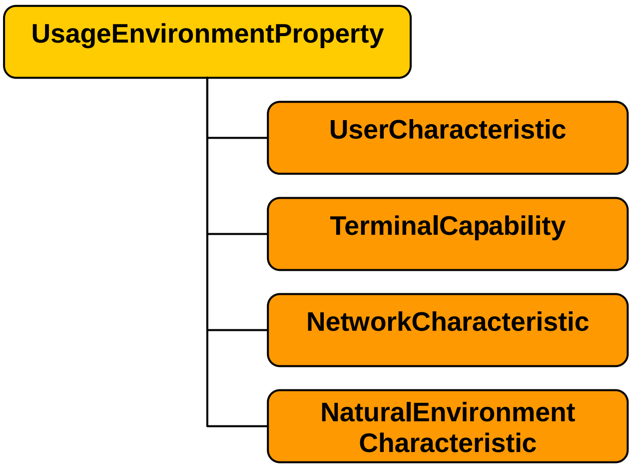


Figure 17. Usage Environment description tools

4.1.3. Usage environment description tools

The Usage Environment Description tools are supposed to be the most frequently used tools in UMA systems. They provide descriptive information about various dimensions of the usage environment, which originate from Users, to accommodate, for example, the adaptation of Digital Items for transmission, storage and consumption. The usage environment includes the description of User Characteristics, terminal capabilities, network characteristics and natural environment characteristics as shown in Figure 17. The details of each description are presented in the next 4 subsections, from 4.1.3.1 to 4.1.3.4.

4.1.3.1. User characteristics

The User Characteristic tools include the content preferences, presentation preferences, accessibility, mobility and destination (Figure 18).
The information about users is described using UserInfo tools:

- UserInfo: UserInfo specifies general information about Users such as name and contact information. A User can be a person, a group of persons, or an organization.

To describe user preferences, UsagePreferences and UsageHistory tools are used. Both of them are derived from MPEG-7 MDS standard.

- UsagePreferences: The UsagePreferences is a tool for describing the preferences of a User related to the type and content of Digital Items. Its detail is described in section 4.2.1.1. For instance, the Usage Preferences can express several preferred genres, namely sports, entertainment, and movies. Such information can be used, for example, by a service provider to personalize the set of Digital Items to be delivered to the User. It can also be used by an agent of the User to automatically filter Digital Items that are broadcast.

- UsageHistory: The UsageHistory descriptor describes the history of actions on Digital Items by a User. As such, it describes the preferences of a User indirectly. These preferences could then be used for the adaptation of Digital Items. The semantics is specified in MPEG-7 MDS part. For example, the consumption history of a User during a particular 6 hour time period (called ObservationPeriod) can be expressed using Usage History. In particular, the items User has played are each identified by a unique identifier. The identifier can be used to identify and/or locate the content descriptions corresponding to each item. The combined information can be used, for example, by a service provider or by a personal agent to infer the preferences of the User, and subsequently provide preferred Digital Items to the User.

The presentation preferences are designed for adapting to the user how to present the content in terms of audio (AudioPresentationPreferences), visual (DisplayPresentation Preferences) and graphics (GraphicsPresentationPreferences). They also can describe preferred resource conversion preferences, priority of presentation and user's attention on multimedia segments (objects, regions, video segments, etc). They can be used either by the server or the terminal to adapt the presentation of the User.

- AudioPresentationPreferences: the preferences of a User regarding the presentation or rendering of audio resources. It represent the audio related preferences to the user, for example, volume, equalizer or preferred audible frequency range. For example, VolumeControl could express that the User has the preference to hear music very loud. The adaptation engine may scale the audio signal to match this preference. FrequencyEqualizer represents the preference of a User to specific frequencies. AudibleFrequencyRange represents the preferred audible frequency range in Hz. SoniferousSpeed could help Users with an auditory impairment to listen to fast speech, User who is studying a foreign language, Users singing in Karaoke to control the playback speed.

- DisplayPresentationPreferences: This specifies the preferences of a User regarding the presentation or rendering of images and videos, such as preferred color, color temperature, brightness, saturation and contrast. An application may convert images so that the resulting images satisfy the User preference for color.

- GraphicsPresentationPreferences: This specifies preferences related to graphics media, such as the preferred degradation of geometry, texture and animation for graphics.

- ConversionPreference: This is a preference to guide the conversion of Resources. For example, in case the User wants to apply generally some conversion rules to video Resources, where it is most desired that the videos be retained if possible (i.e., order of video-to-video is 1), and if videos must be converted, they should be converted to audios first (order of video-to-audio is 2). If the resources again must be converted, it may ultimately be converted to image or text.

- PresentationPriorityPreference: This is a tool to let the User have choices on the presentation qualities of different resources at the output of the content adaptation process.

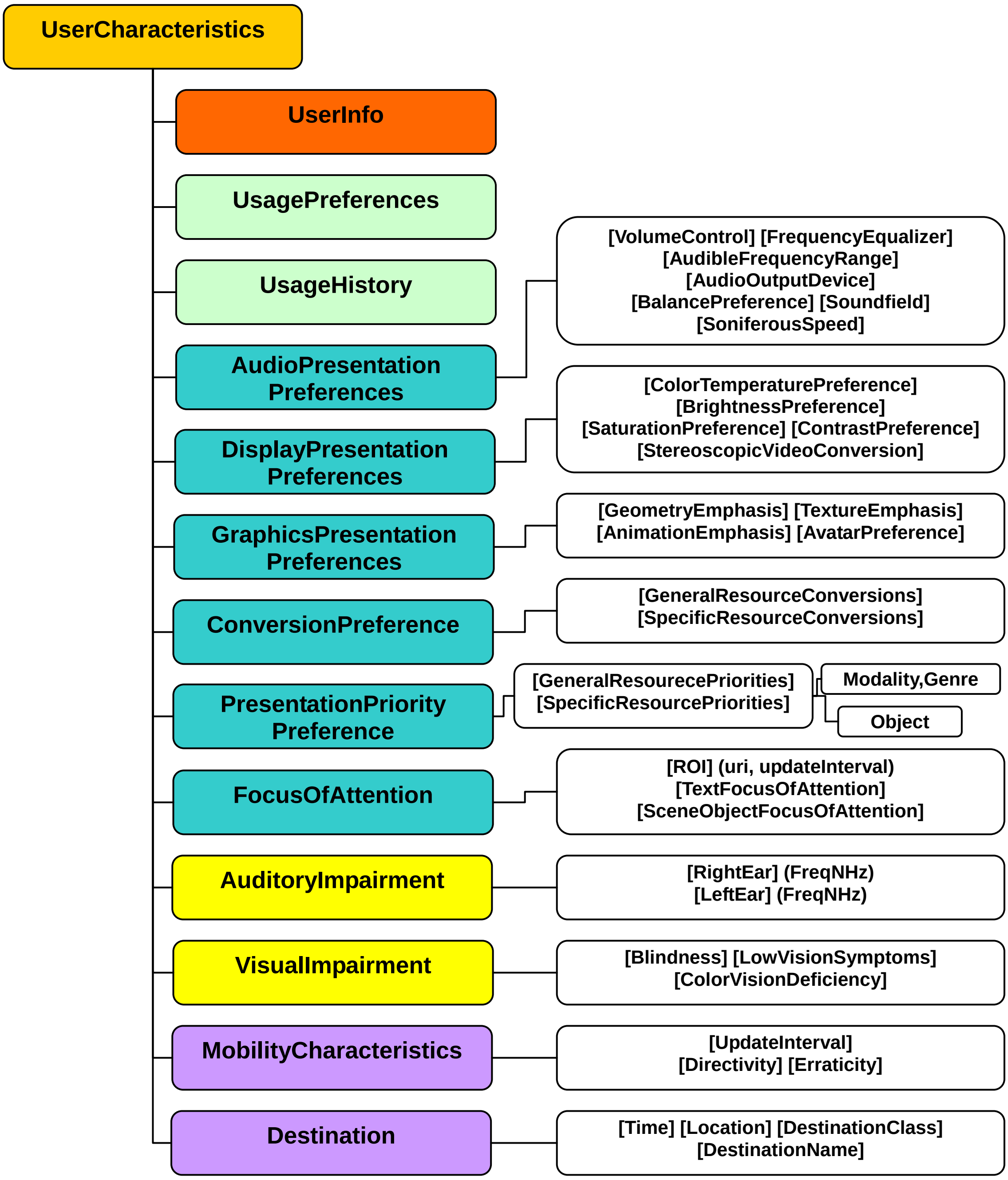


Figure 18. User Characteristic description tools

An example is when a User accesses a Miss World website and the User is interested in images. For this case, the User gives a high priority for image resources. The result is that images will be adapted with higher quality than usual (i.e. without having the User's priorities). Yet, other resources will be of course degraded because the total bandwidth constraint is likely fixed.

- FocusOfAttention: This specifies the User's preferences related to multimedia segments (eg. region of interest (ROI)). Specifically, the focus of attention in a given resource such as audio, visual, audio-visual and/or text contents. For example, FocusOfAttention tools could help adapting the contents to a User that wants to watch a particular player, i.e., region of interest

(ROI), in the video of a basketball game, or any other interested segments.

Two tools are specified to describe the characteristics of a User's difficulty in hearing and seeing, AuditoryImpairment tools and VisualImpairment tools.

- AuditoryImpairment: This is used to describe the characteristics of a particular User's auditory deficiency. The description of these measurements can help an audio resource adaptation engine to provide an improved quality of audio by compensating the hearing loss in one ear.

- VisualImpairment: VisualImpairment covers a wide range of conditions. The various forms of visual impairment include difficulty to read the fine print, low vision that cannot be corrected by standard glasses, total blindness, color vision deficiency, i.e., the inability to recognize certain colors. The low vision conditions due to their wide variety, are described by the User's symptoms, but the names of conditions are not described.

To describe the User's position or movement, the MobilityCharacteristics tools and the Destination tool are specified.

- MobilityCharacteristics: This tool describes the mobility characteristics of a User (UpdateInterval, Directivity, Erraticity). It is mainly used to assume the User's transportation means or actions so that application service provider (ASP) can provide the best service adaptive to the mobile profile.

- Destination: a tool for describing the destination of a User. (Time, Location, DestinationClass, FreeClass, StereotypedClass, DestinationName)

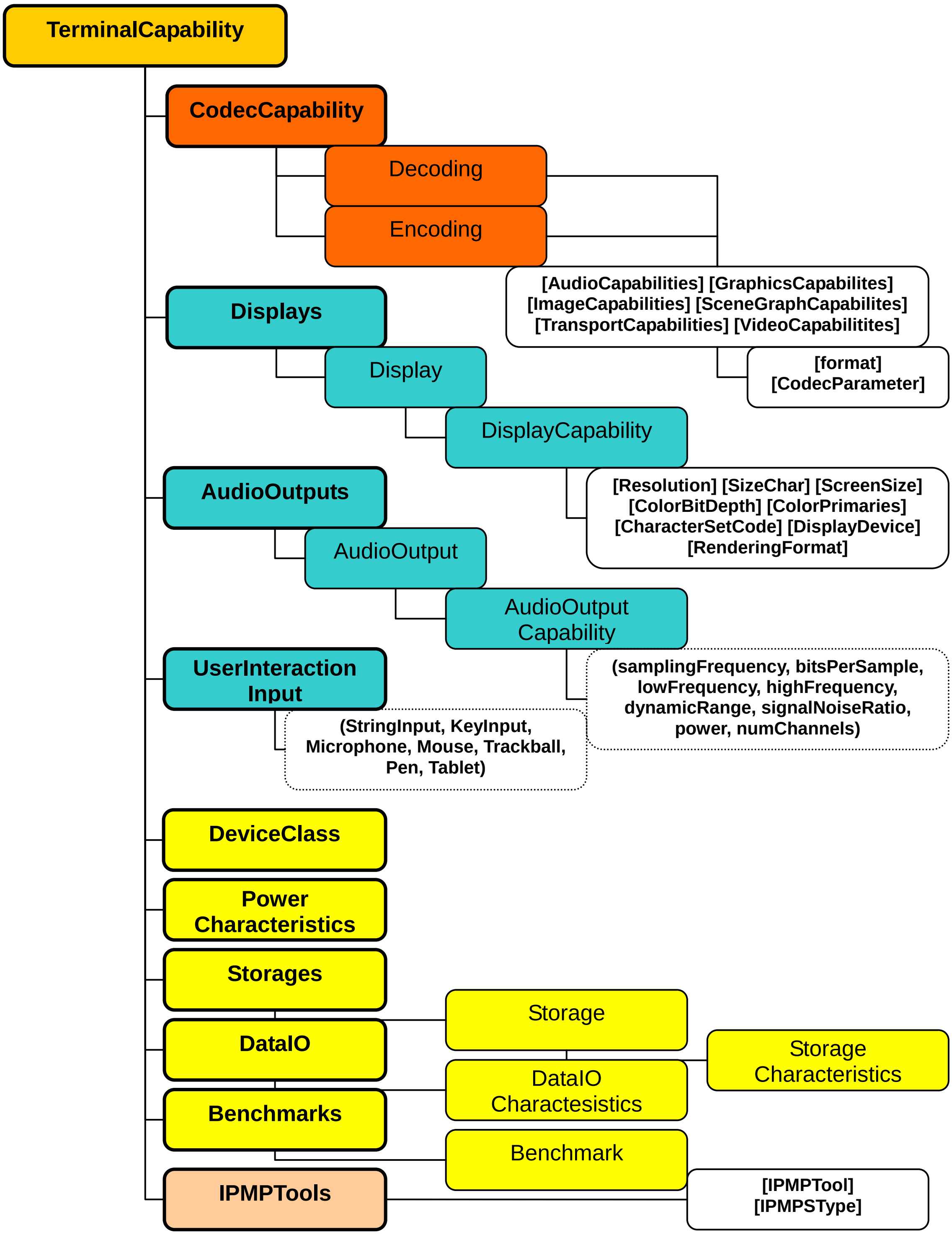


Figure 19. Terminal Capability description tools.

4.1.3.2. Terminal capabilities

The Terminal capability description tools include terminal capabilities in terms of coding and decoding capabilities, device properties and input-output capabilities. Figure 19 shows the Terminal Capability description tools. These tools are quite useful to assume for example at which format, at which bit-rate the content could be playable on the user's terminal.

CodecCapabilities tool are designed to describe the coding and decoding capabilities of the terminal.

- CodecCapabilities: Tool for describing the encoding and decoding capabilities of the terminal. Codecs for audio, graphics, image, video, scene graph and transport formats are included, and their codec parameters, such as buffer size, bitrate, memory bandwidth, vertex processing rate, fill rate of a graphics codec, can also be described.

- Decoding: Describes the decoding capability of the terminal.

- Encoding: Describes the encoding capability of the terminal.

Input-output capabilities of displays and audio outputs can also be described using Display tools and AudioOutput tools, respectively.

- DisplayCapability: Tools for describing the capability of a single display or multiple displays such as resolution, screen size, color bit depth.

- AudioOutputs: Tools for describing the capabilities and properties of a single audio output or multiple audio outputs, such as sampling frequency and dynamic range.

- UserInteractioninput: Tools for describing the User interaction input support that is available on a particular device. With such information available, an adaptation engine could modify the means by which a User would interact with resources contained in a multimedia presentation.

The device properties including device class, power, storage, data I/O and benchmarks .

- DeviceClass: Describes the type of terminal such as PC, PDA, STB, Printer, Mobile phone, digital still/video camera, audio player, TV, gateway and router.

- PowerCharacteristics: Describes the average ampere consumption, remaining capacity of a battery, and time remaining of a battery.

- Storages: Describes the storage characteristics of terminal(s). It consists of the input/output transfer rate of the storage device, storage size and if the storage device can be written to or not.

- DataIOs: Specifies data input-output characteristics of the terminal(s). It consists of the width of the bus, transfer speed that the bus is capable, and the maximum and current number of devices supported by the bus.

- Benchmarks: Speficies benchmarks and their results. It consists of benchmarks of device, CPU and graphics performance.

- IPMPTools: Specifies the characteristics of Intellectual Property Management and Protection (IPMP) tools of the terminal to facilitate the adaptation of the protected Digital Items. IPMP tools are modules that perform one or more IPMP function, such as authentification, decryption, watermarking, etc..

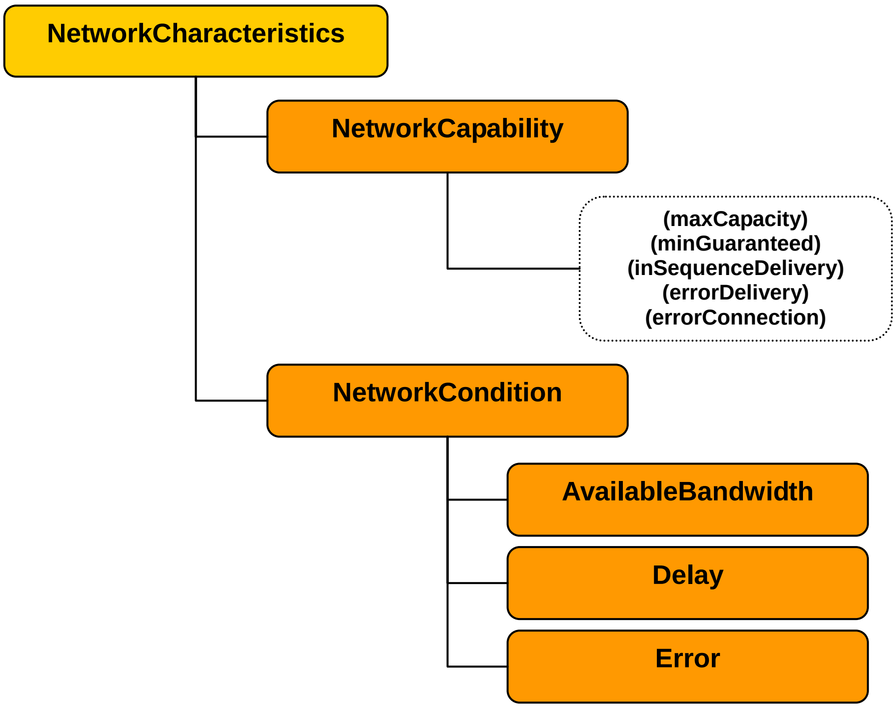


Figure 20. Network Characteristics description tools.

4.1.3.3. Network characteristics

The description of Network Characteristics consists of network capabilities and conditions. They include available bandwidth, delay and error characteristics. Figure 20 illustrates the Network Characteristics description tools.

Network capabilities tools specify the static capabilities of a network.

- maxCapacity: maximum bandwidth capacity of a network in bits/sec.

- minGuaranteed: minimum guaranteed bandwidth of a network in bits/sec.

- inSequenceDelivery: Describes the capability of a network to provide in-sequence delivery of data units.

- errorDelivery: Describes whether data units containing errors are delivered or dropped by the network.

- errorCorrection: Describes whether data units containing errors are corrected or not by the network.

NetworkCondition tools specify the dynamic conditions of a network.

- AvailableBandwith: describes the available bandwidth of a network.

```
<DIA>
  <Description xsi:type="UsageEnvironmentType">
    <UsageEnvironment xsi:type="NetworksType">
      <Network xsi:type="NetworkType">
        <NetworkCharacteristic xsi:type="NetworkCapabilityType"
        maxCapacity="256000" minGuaranteed="32000"/>
        <NetworkCharacteristic xsi:type="NetworkConditionType">
          <AvailableBandwidth maximum="256000" average="80000" interval="330"/>
          <Delay packetTwoWay="330" delayVariation="66"/>
          <Error packetLossRate="0.05"/>
        </NetworkCharacteristic>
      </Network>
    </UsageEnvironment>
  </Description>
</DIA>
```

Figure 21. Example of network characteristic description

- Delay: Describes the delay characteristics of a network, such as one-way or round-trip packet delay and the difference between the one-way packet delay of two successive packets.

- Error: Describes the error characteristics of a network. It includes packet loss rate and bit-error rate on a specified channel.

Network characteristics examples.

The following example in Figure 21 describes a network that is characterized by a maximum capacity of 256 kbps and a minimum guaranteed bandwidth of 32 kbps. Over amn interval of 330 milliseconds, this description indicates that the maximum bandwidth achieved was 256 kbps and the average over that time was 80 kbps. Other attributes for the delay and error characteristics are described as well.

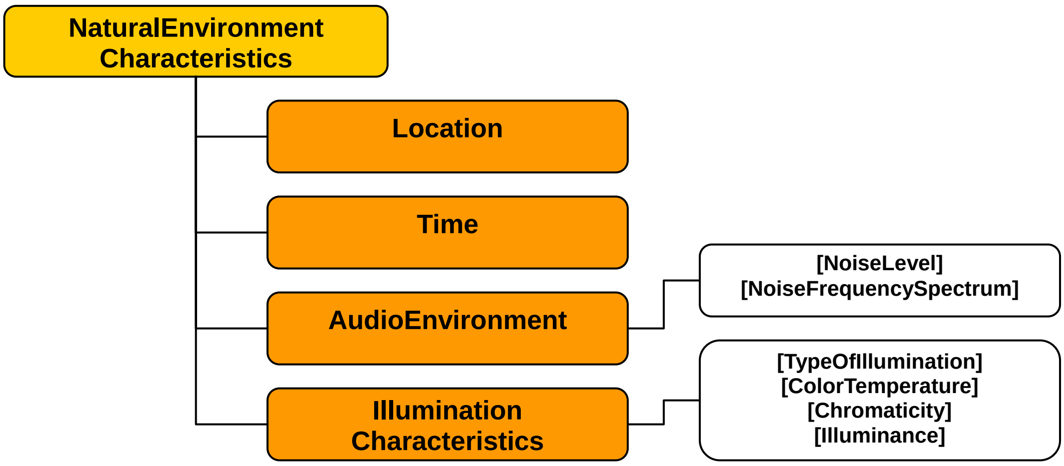


Figure 22. Natural Environment Characteristics description tools.

4.1.3.4. Natural environment characteristics

The description of natural environment characteristics consists of the location and time of usage of a Digital Item, as well as audio-visual characteristics of the natural usage environment. Figure 22 shows the natural environment characteristics description tools.

- Location: Describes the location of the usage of a Digital Item. For example, Location tools can express a precise geographic position by giving specific coordinates for latitude, longitude and altitude.

- Time: Describes the time of the usage of a Digital Item.

- AudioEnvironment: Describes the natural audio environment of a particular User. It consists of the noise level and the noise frequency spectrum. Both of them can be acquired by processing noisy signal input from a microphone of the User's terminal. This description can be used by an adaptation engine for automatically adjusting the audio signal level to the terminal. The adaptation engine may reside in the terminal and responds automatically to the changing noise level of the environment.

- IlluminationCharacteristics: Describes the illumination characteristics of the natural environment, which includes type of illumination, color temperature, chromaticity and illuminance. The overall illumination around a display device affects the perceived color of images on the display device and is a factor causing distortion or variation of perceived color. With the information on the type and illuminance of the overall illumination, such affects on the perceived color can be estimated. For example, such information can be used to estimate the chromatic adaptation of perceived color on chromaticity coordinates. By compensating the estimated distortion, actual distortion caused by the overall illumination can be lessen or removed.

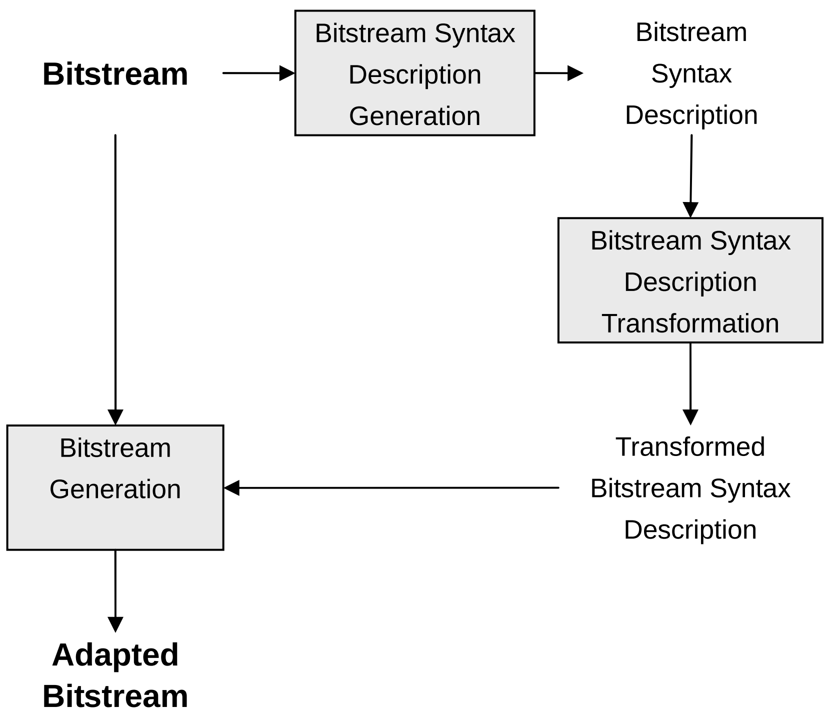


Figure 23. Adaptation architecture.

4.1.4. Bitstream syntax description

A BSD describes the syntax (high level structure) of a binary media resource. Using such a description, a Digital Item resource adaptation engine can transform the bitstream and the corresponding description using editing-style operations such as data truncation and simple modifications. It consists of two main technologies, Bistream Syntax Description Language (BSDL) and Generic Bitstream Syntax (gBS). BSDL is an XML schema based language to design specific bitstream syntax schemas for **particular media formats**. gBS schema is a generic schema enabling the construction of **resource format independent** bitstream syntax descriptions.

Figure 23 illustrates the architecture of a resource adaptation step. The Bitstream Syntax Description generator parses a bitstream described by a Bitstream Syntax Schema and generates its Bitstream Syntax Description. The bitstream and its Bitstream Syntax Description are subject to the adaptation. An adaptation engine is assumed to determine the optimal adaptation for the media resource given the constraints as provided by the DIA Descriptions. Based on that decision, if the resource is not pre-stored but needs to be derived by adapting an existing resource, then one (or several) Bitstream Syntax Description Transformations is (are) selected to be applied to the input description. The result of these transformations is a Transformed Bitstream Syntax Description which is the base for the generation of the adapted bitstream. The Bitstream Syntax Description may be the instance of either a specific Bitstream Syntax Schema or the normative generic Bitstream Syntax Schema.

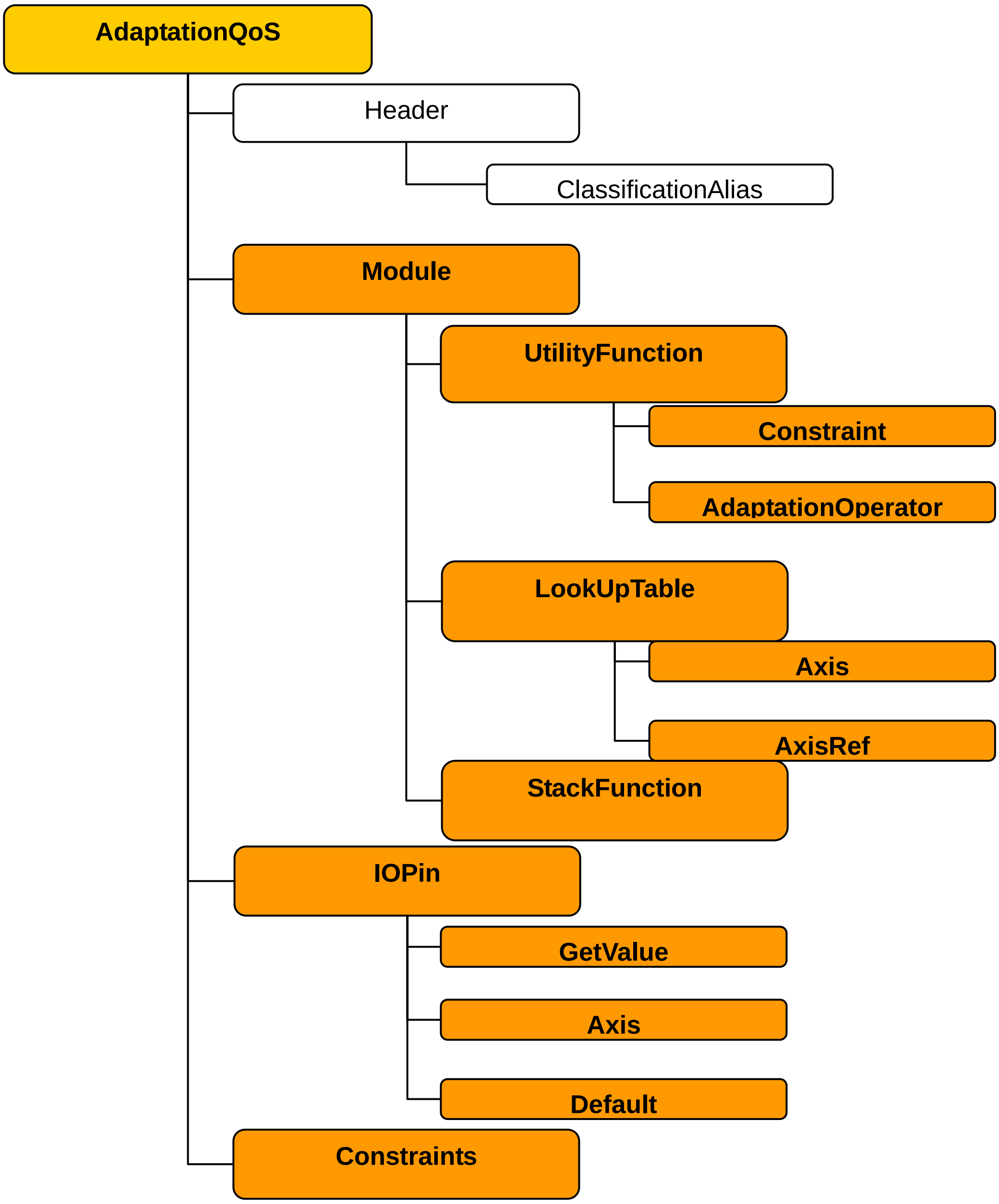


Figure 24. Terminal and Network quality of service description tools

4.1.5. Terminal and network quality of service

Terminal and network quality of service (QoS) addresses the problem of media resource adaptation to constraints imposed by terminals and/or networks for QoS management. The AdaptationQoS descriptor specifies the relationship between constraints, feasible adaptation operations satisfying these constraints, and associated utilities (qualities). Therefore, the AdaptationQoS tool lets an adaptation engine know what adaptation operations are feasible for satisfying the given constraints and the quality resulting from each adaptation. In this way, terminal and network QoS management is efficiently achieved by adaptation of media resources to constraints.

In general, the AdaptationQoS description is generated in a media resource server and is delivered along with the associated media resource to an adaptation engine located at a network proxy or a terminal. The generation of the AdaptationQoS description can be done for each media resource stored in a server in advance in the case of on-demand applications. In the case of streaming of live events, the description could be generated by a prediction-based approach in real-time.

The main constraints in media resource adaptation are bandwidth and computation time. Adaptation Methods include selection of frame dropping and/or coefficient dropping, requantization, MPEG-4 fine Granular Scalability(FGS), wavelet Reduction and spatial size reduction.

Figure 24 shows Terminal and Network quality of service description tools.

Header: Describes an optional list of Classification Scheme aliases.

Module: Describes a list of AdaptationQoS modules which can be linked together.

- UtilityFunction: Tool for describing the mapping relations among feasible adaptation operation(s), associated qualities and given constraint(s) in a list format, used for sparse, discrete data representation. Linear interpolation is assumed between constraint points.

- LookUpTable: additional multi-dimensional sets of data to support more elaborate adaptation scenarios.

- StackFunction: tool for describing the data in numerical function format.

IOPin: Tool for describing the globally declared interface of the modules for linking modules and for external referencing.

Constraints: Describes an optional list of constraints as defined in the Universal Constraints Descriptor.

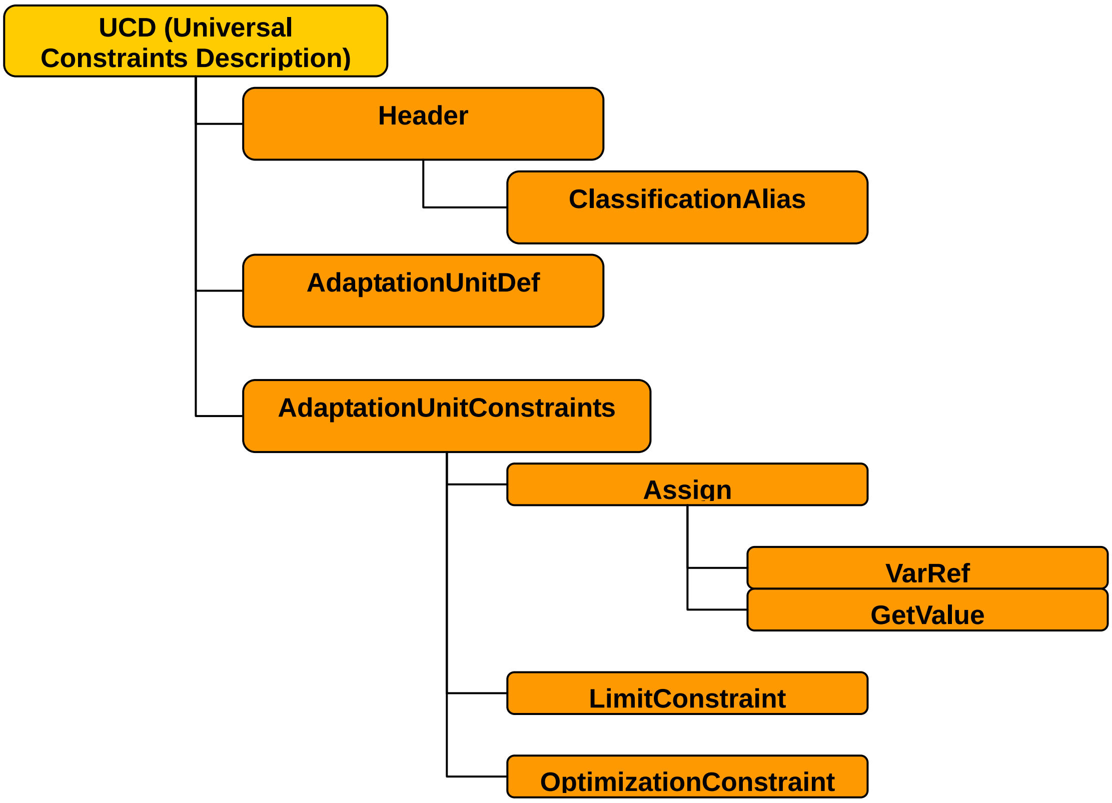


Figure 25. Universal constraints description tools

4.1.6. Universal constraints description tools

This subsection specifies the Universal Constraints Description (UCD) tool for describing constraints for adaptation. The UCD can be provided both from the consumer to an adaptation engine, and from the content provider in conjunction with AdaptationQoS. In the former case, the UCD supplements the information in the Usage Environment Descriptors and also converts it into a semantics free form for format-independent decision-making. In the latter case, the UCD allows content providers to specify provider side constraints that must be satisfied for any adaptation of a resource.

Header: Describes an optional list of Classification Scheme aliases.

AdaptationUnitDef: Describes the adaptation unit axis. If used in conjunction with the AdaptationQoS, this corresponds to an IOPin whose axis definitions are expected to be present in the AdaptationQoS.

AdaptationUnitConstraints: Describes constraints for a single adaptation unit (GOP, frames, ROI etc.). There can be an unbounded number of AdaptationUnitConstraints elements, one for each adaptation unit. If the number of such elements is less than the actual number of adaptation units expected based on the AdaptationUnitDef element, then the last AdaptationUnitConstraints element is applied to all remaining adaptation units.

- **AssignType:** Describes an assignment. VarRef describes the target of the assignment, and GetValue describes the value to be assigned.

- **LimitConstraintType:** Describes a constraint by means of a metric represented in stack function form, which must evaulate to Boolean true, for any valid adaptation decision.

- **OptimizationConstraintsType**: Describes a metric to optimize represented in stack function form.

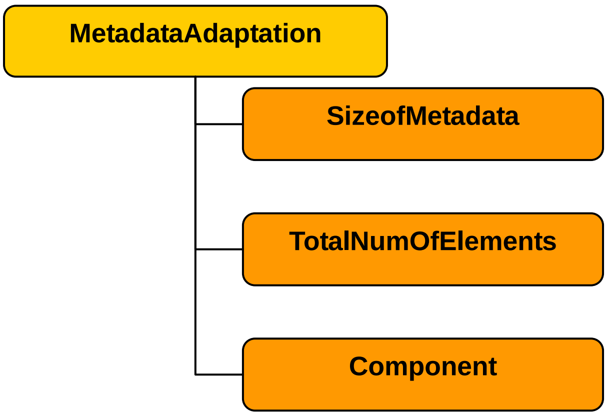


Figure 26. Metadata Adaptavility

4.1.7. Metadata adaptability

Metadata Adaptation tools describe adaptation hint information pertaining to metadata within a digital item. This information is a set of syntactical elements with prior knowledge about the metadata that is useful for reducing the complexity of the metadata adaptation process. Some possible examples are resizing of a description (scaling) and integration of two descriptions.

Figure 26 illustrates the metadata adaptability tools. Its semantics are in the following.

- SizeOfMetadata: Describes the size of the metadata description in bytes.

- TotalNumOfElements: Describes the optional total number of words (i.e., XML elements and values) that are included in a metadata description.

- Component: Describes a target element in a metadata description.

4.2. MPEG-7 tools for adaptation

There are four tools for adaptation in MPEG-7 [MPEG7MDS]; tools for personalization, variation, summarization and transcoding hints [PBeek03].

For (personalized) content selection, one or more of 1) multimedia content descriptions, 2) user preference descriptions, and 3) content usage history is used. Multimedia content descriptions allows the user to query the system and to search for desired content based on attributes like author, title, genre, language, keywords, etc. Information filtering, which utilizes a user profile to capture long-term preferences, and collaborative filtering, which applies to communities of users that share their explicit opinions or ratings of content items, is introduced to enable automatic filtering or recommendation services. MPEG-7 UsageHistory DS and UserPreference DS support these filtering functionalities [Vetro03]. The details of them are described in clause 4.2.1

To deliver the selected content to the user, the variation which adapts to the client terminal capabilities or user preferences should be selected or created (transcoded, summarized, etc.). MPEG-7 Variation Tools enables to describe a single content using various spatial and temporal resolution, quality, coding format, bit rate, color detail, length and modalities (video/image/audio/text). MPEG-7 media transcoding hints allow content servers, proxies, or gateways to adapt AV contents to different network conditions, user and publisher preferences, and capabilities of terminal devices with limited resources. Transcoding hints can be used for complexity reduction as well as for quality improvement in the transcoding process. MPEG-7 summary descriptions defines the summary content, how it relates to the original content, and how an actual summary of the original content can be composed from these and presented to the user.

This clause presents personalization and variation tools in MPEG-7, which are the most relevant for multimedia content adaptation.

4.2.1. MPEG-7 tools for personalization / customization of multimedia contents

The key concepts used in this subsection are illustrated in Figure 27. A user interacts with multimedia content by using a multimedia system. The multimedia system is used to find multimedia content, e.g. by searching or filtering, and to consume multimedia content, e.g., by viewing or listening. Descriptions of the multimedia content are provided to the system to enable efficient searching, filtering and browsing. Descriptions of the user's preferences are also provided to the system to enable personalized searching, filtering and browsing of multimedia content. The multimedia system may also generate a usage history description based on a history of the user's interactions with the multimedia content. The usage history descriptions may be used directly for personalized searching, filtering and browsing, or may be mapped to a description of the user's preferences. Both user preferences descriptions and usage history descriptions may be exchanged with third parties (e.g. service providers) or with other devices.

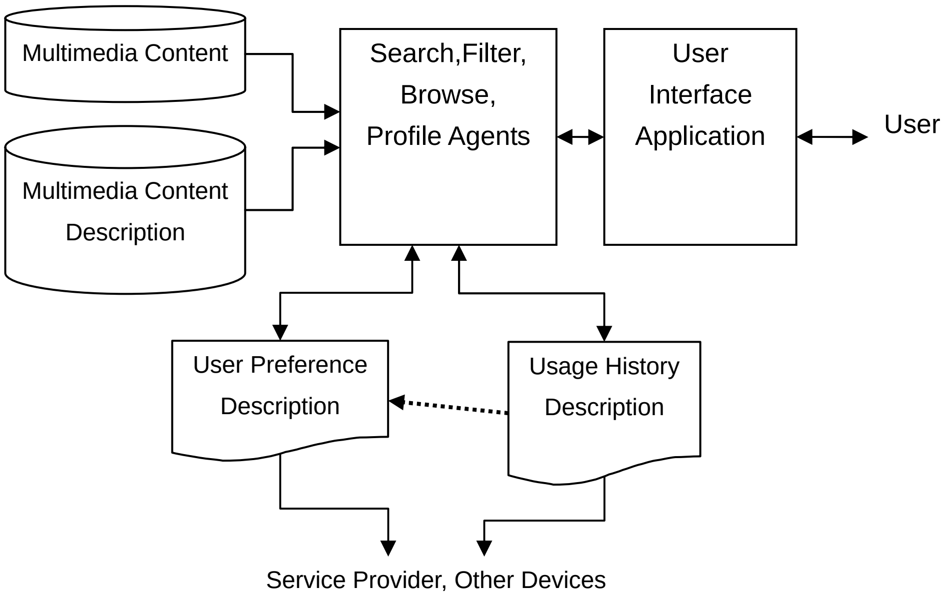


Figure 27. Overview of an interactive multimedia content personalization system

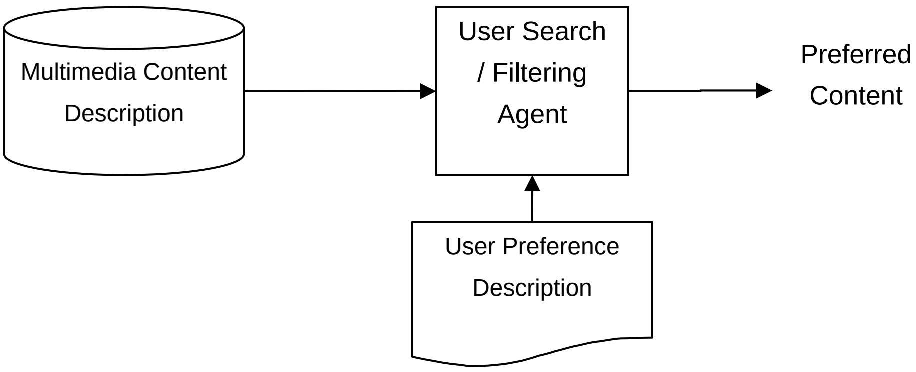


Figure 28. Overview of usage model for user Preference and content descriptions.

4.2.1.1. User preference description tools

This subsection specifies tools for describing user's preferences pertaining to consumption of multimedia content. User preference descriptions can be correlated with content descriptions to find and consume desired content. Correspondence between user preferences and content descriptions facilitates accurate and efficient personalization of content access and content consumption.

User preferences descriptions can be utilized by consumers (or their agents) for accessing and consuming multimedia content that fits their personal preferences. A generic usage model is depicted in Figure 28, where a user agent takes content descriptions and user preferences as input and generates a filtered output specifying the content item(s) that fit personal preferences. The descriptions of the user's preferences are used to find preferred multimedia content and to present preferred views of the content automatically. In specific applications, the output may include identifiers or media locators of preferred multimedia content, or a summary of a multimedia program where the type of the summary satisfies user's summary preferences.

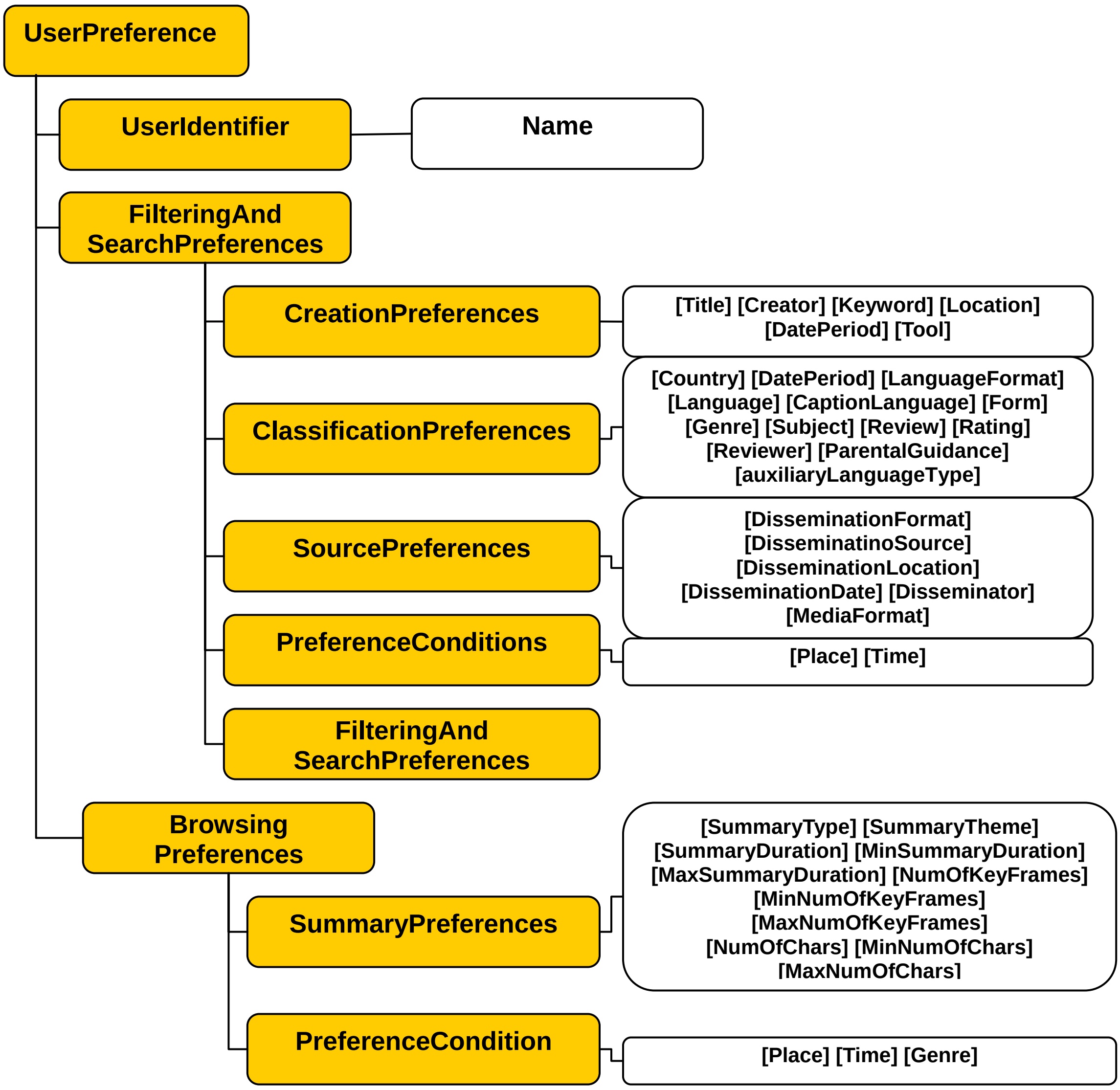


Figure 29. User Preference Description Scheme

For example, a particular user may indicate a preference for movies of a certain genre, in which case movies of other genres may be filtered out by a user agent or ranked lower in an ordered list of multimedia content. Another example is a user who prefers to view only the goals of a soccer match, while yet another user may prefer a 30-minute highlight summary of the entire match.

Figure 29 illustrates the structure of the UserPreferences Description Scheme (DS). The UserPreferences DS is used to describe the user's preferences pertaining to consumption of multimedia content, in particular, filtering, searching and browsing of multimedia content. The UserPreferences DS contains FilteringAndSearchPreferences and BrowsingPreferences, and contains an attribute indicating whether the user's preferences may be updated automatically.

UserIdentifier identifies a particular set of user preferences and distinguishes it from other sets of user preferences.

FilteringAndSearchPreferences describes preferred multimedia content, in terms of attributes related to the creation, classification and source of the content. Preferred content may be determined by matching individual components or combinations of components of a FilteringAndSearchPreferences description against descriptions of multimedia content. First level preference components are CreationPreferences, ClassificationPreferences and SourcePreferences. Each of these elements in turn contains second level preference components. A FilteringAndSearchPreferences element may optionally contain other FilteringAndSearchPreferences elements as its children, to specify hierarchically structured preferences. In this case, the filtering and search preferences of the children nodes apply on the condition that the preferences contained in their ancestor nodes are satisfied by matching multimedia content.

- CreationPreferences: The CreationPreferences DS is used to describe user preferences related to the creation of the multimedia content, such as preference on a particular title, or a preferred actor, or preferred period of time within which the content was created or preferred place where the content was created, or preferred tools used in the creation. The user may also include keywords to describe preferred multimedia content.

- The ClassificationPreferences DS is used to describe user preferences related to classification of the multimedia content, e.g., preferred genre and form of the content, preferred country and time the content was released, or preferred language of the spoken content or captions.

- The SourcePreferences DS is used to describe user preferences related to the source of the multimedia content, such as a preferred dissemination medium, or a preferred distributor or publisher, or preferred place and date where and when it is made available for consumption, or a preferred format for the media.

The BrowsingPreferences DS is used to describe user preferences pertaining to navigation of and access to content. In particular, a user may express preferences on the type and content of summaries of multimedia content. These preferences may be conditioned on certain times and locations, and type of multimedia content in terms of genre.

- The SummaryPreferences is used to describe user preferences for nonlinear navigation and access to the multimedia content, in particular with regard to summarization. Users can specify their preferences for multiple alternative summaries of multimedia content that fit best to their desire and constraints.

- The PreferenceCondition describes the usage condition(s) for a particular browsing preference description, in terms of time and place, and genre of the multimedia content.

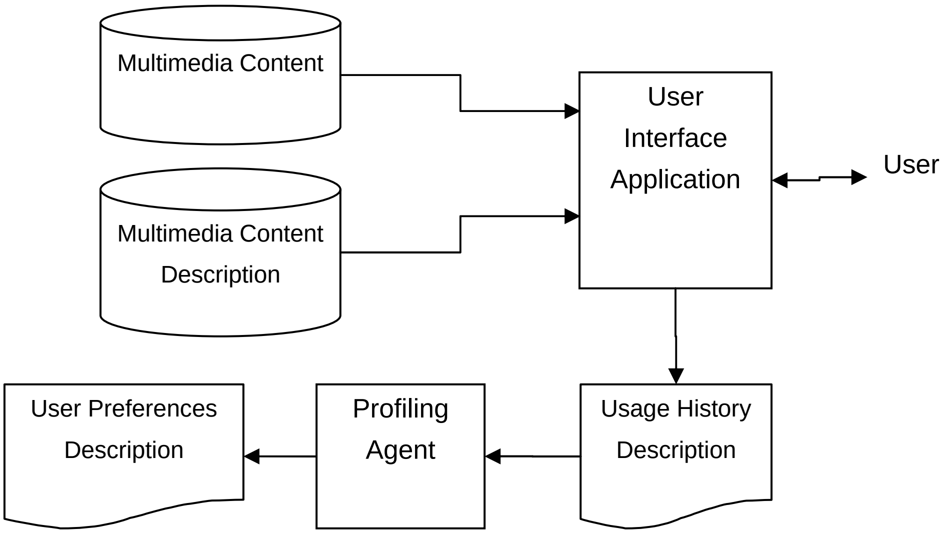


Figure 30. Overview of usage history descriptions and their applications.

4.2.1.2. Usage history description tools

This subsection specifies tools for describing a history of the actions that consumers of multimedia content have carried out over a specified period of time. Usage history descriptions can be exchanged between consumers, their agents, content providers, and devices. User agents and content providers may, for example, use these descriptions of a user's usage history to determine the user's preferences with regard to multimedia content.

A generic context diagram is provided in the Figure 30, showing a user interface application that takes into account the user’s interactions with the multimedia content as well as multimedia content descriptions, and produces as output organized descriptions of the user’s content consumption history. The red arrows indicate possible uses of this usage history information by other agents, devices or services (if permitted by the user).

The UsageHistory DS structure is shown in Figure 31. The UsageHistory DS is used to describe a set of UserActionHistory elements, each with its own observation period. The UserActionHistory DS is used to describe a set of UserActionList elements, each of which contains UserAction elements of a single type. The UsageHistory DS can be used to form a compact description of user action information and related statistics. The usage history descriptions consist of action type-specific lists that include identifiers of the programs associated with each action. Also, the time of user actions can be indicated, as well as the time-extent of the multimedia content that was consumed. Finally, a reference to content descriptions may optionally be added to the description of each user action, to scope the parts of the content the user action is associated with.

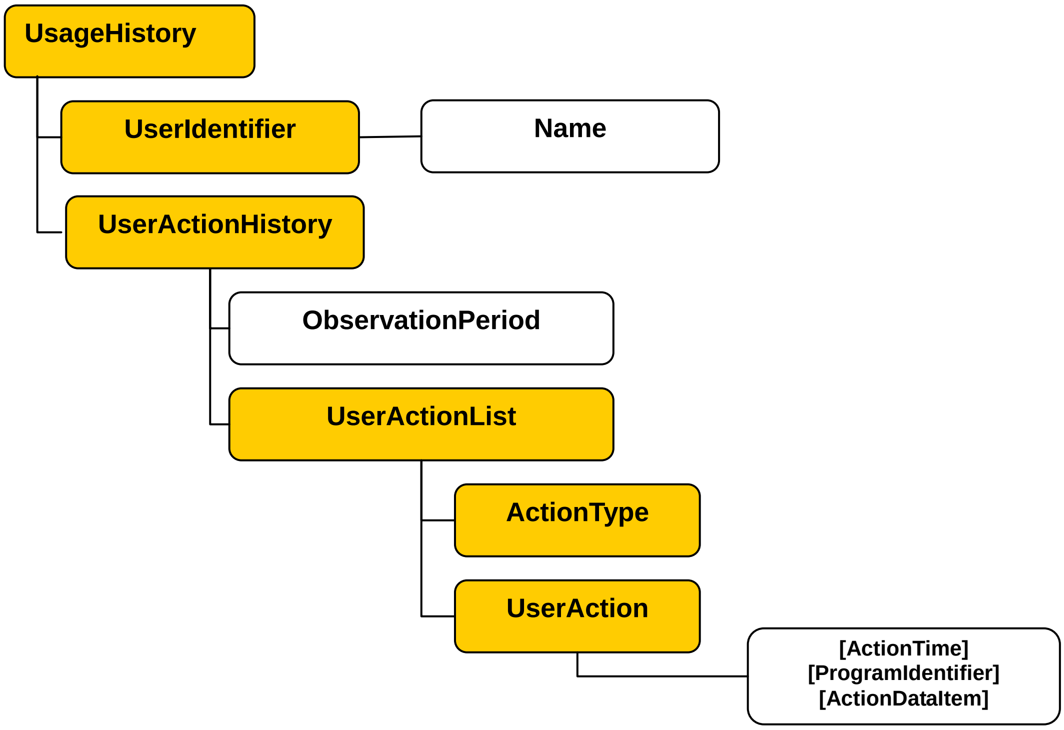


Figure 31. Usage History Description Scheme

4.2.2. MPEG-7 tools for describing variations of multimedia contents.

4.2.2.1. Variation tools

Variation tools in MPEG-7 enable to describe the associations or relationships between different variations of multimedia content. The Variation DS supports content management by tracking the variations of multimedia content that result from various types of multimedia processing such as summarization, translation, reduction, revision, transcoding and so forth. The Variation DS also supports Universal Multimedia Access by allowing the selection of the most appropriate variation of the multimedia content for the specific capabilities of the terminal devices, network conditions or user preferences. Figure 7 in clause 3.2.2 illustrates a set of variations of multimedia content.

This tool can specify the type of association or relationship of the variation multimedia content with the source content. The different types of relationships are given as follows:

- Summarization: summarizes the important information of the source content in order to support efficient browsing and navigation. Examples of summaries include those defined in the Summary DS such as hierarchical summaries and sequential summaries.
- Abstraction: presents the salient points of a multimedia content. An abstract differs from a summary in that the abstract is separately authored while a summary is derived from the multimedia content.
- Extraction: extracts multimedia content from a multimedia content. Example extracts include key frames selected from video, audio-band and voice excerpts from audio content, paragraphs and key terms from text, and regions, segments, objects, and events from audio

and video programs.

- ModalityTranslation: involves the conversion from one multimedia content modality (image, video, text, audio, graphics) to another. Examples of modality translation include text-to-speech (TTS), speech-to-text (speech recognition), video-to-image (video mosaicing), image-to-text (video-text or embedded caption recognition), and graphics-to-image (graphics rendering).
- LanguageTranslation: involves the conversion of text or speech associated with a multimedia content from one language to another, such as Spanish to English.
- ColorReduction: involves the reduction of the color detail of visual content such as an image or video. Examples include the reduction of a 24-bit RGB color image to 8-bit grayscale.
- SpatialReduction: involves the reduction in the spatial size of visual content such as an image or video. Examples include the spatial reduction of frames in a video from spatial size 640x480 to 320x240 (width x height).
- TemporalReduction: involves the reduction in the time duration of audio or video content. Examples include the temporal reduction of a video from 1 hour to 15 minutes.
- SamplingReduction: involves the reduction in the temporal sampling period of audio or video content without shortening of the overall playback time. Examples include the frame rate reduction of video from 30 frames/sec to 15 frames/sec.
- RateReduction: involves the reduction in the temporal data rate of audio or video content. Examples include the rate reduction of MP3 audio from 128 Kbps to 96 Kbps.
- QualityReduction: involves the reduction in the detail or perceived quality of the multimedia content. Examples include the reduction of quality associated with the quantization of images from 256 levels of intensity to 32 levels of intensity.
- Compression: involves the reduction in the amount of data used in representing multimedia content. Examples include the lossy compression of a 1MB image to 256KB.
- Scaling: involves selecting a representation of a multimedia content from one of the levels in a scalable data representation. Examples include the scaled representation of an image from the coarse level of a multi-resolution image pyramid.
- Revision: indicates that a multimedia content has been revised in some way, such as through editing or post-processing, to produce the variation.
- Substitution: indicates that one multimedia content can be used as a substitute for another, without specifying any other explicit relationship between the programs. Examples of substitution include "alt" text in HTML, which is text that can substitute for an image that cannot be handled by the terminal device, or an audio track that replaces a chart in a presentation.
- Replay: indicates that the variation content refers to a replay such as a slow-motion replay of the source content.
- AlternativeView: indicates that the variation content provides an alternative view of the source content such as a view from another camera.
- AlternativeMediaProfile: indicates that the Variation refers to an alternative media profile of the source content.

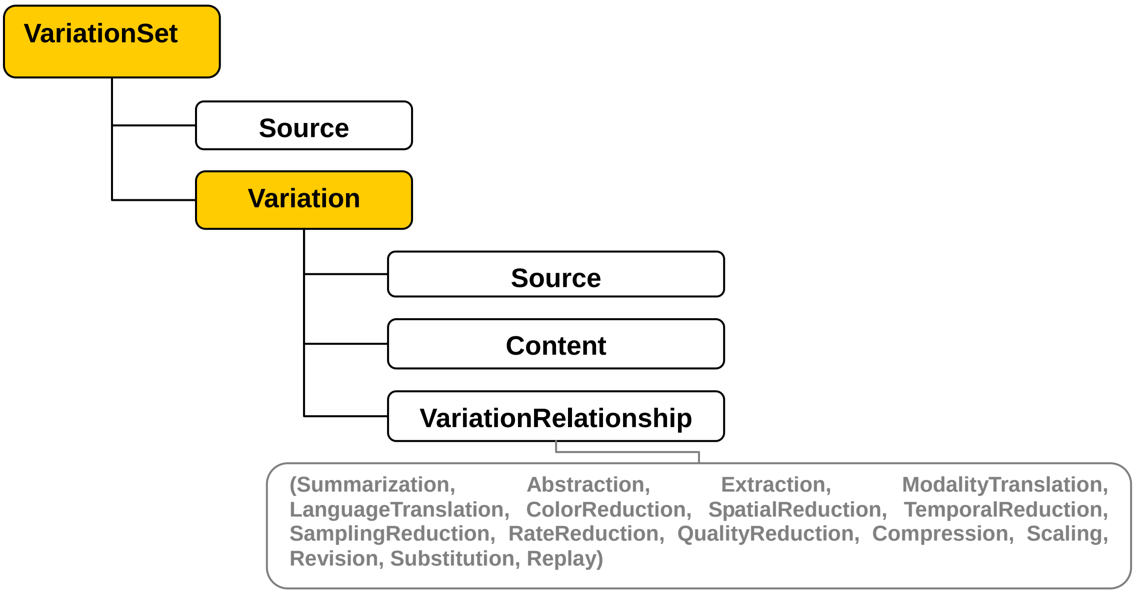


Figure 32. Variation Description

Figure 32 illustrates variation description tools. **VariationSet** describes a set of variations of a multimedia content source. It consists of **Source** element to describe the source multimedia content and **Variation** element to describe a set of variations.

- **Source**: Describes the source multimedia content. Exactly one source shall be specified. If a source is specified within an embedded Variation, then that embedded source shall override the value of this source content for that particular Variation.

- **Variation**: Describes the unbounded set of variations of the source multimedia content. At least one variation shall be specified.

Variation describes the variation multimedia content and its relationship to the source multimedia content.

- **Source**: Describes the source multimedia content. At most one Source shall be specified. In the case that the VariationType is embedded within a VariationSet, then the specification of Source is optional.

- **Content**: Describes the variation multimedia content.

- **VariationRelationship**: Describes the different types of association relationships of the Variation multimedia content with respect to the source multimedia content. Multiple VariationRelationships may be specified.

4.3. Other relevant standards

4.3.1. Usage environment description (CC/PP, UAProf, PSS, Device Indp. WG)

CC/PP (Composite Capability/Preference Profiles) by W3C [CC/PP].

A CC/PP profile is a description of device capabilities and user preferences that can be used to guide the adaptation of content presented to that device. The Resource Description Framework (RDF) is used to create profiles that describe user agent and proxy capabilities and preferences. A CC/PP profile contains 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. The CC/PP vocabulary is a set of identifiers (URIs) used to refer to specific capabilities and preferences, including the types of values to which CC/PP attributes may refer a description of how to introduce new vocabularies and a small client vocabulary covering print and display capabilities. This group was merged into Device Independence Working Group in March 2003.

UAProf (User Agent Profile) by WAP Forum [UAProf].

A User Agent Profile (UAProf) is a description of device capabilities and user preferences that can be used to guide the adaptation of content presented to that device based on CC/PP and RFD. UAProf defines a format to describe device attributes, a core vocabulary defining specific device attributes, a protocol for the client to inform servers about its attributes (extension of HTTP) and Rules determining how servers resolve profiles from the information sent by clients.

PSS (Transparent end-to-end packet switched streaming service) by 3GPP.

The 3GPP PSS [PSS232][PSS233][PSS234] provides a framework for Internet Protocol (IP) based streaming applications in 3G networks. PSS is also based on CC/PP and RDF.

Device Independence Working Group in W3C [DI].

The Device Independence Working Group discusses the challenges that authors commonly face when building web content and applications that can be accessed by users via a wide variety of different devices with different capabilities. The goal is quite related to MPEG-21 DIA as this group also tries to describe the usage environments in a standardized form.

4.3.2. Content metadata and user preference description (TV Anytime Forum)

TV Anytime Forum [TVA].

The TV-Anytime Forum is an association of organizations which seeks to develop specifications to enable audio-visual and other services based on high volume digital storage in consumer platforms (local storage). The TV-Anytime Metadata Specification part has liaison with MPEG-7 and enables describing content used e.g. in Electronic Program Guides (EPG), or in Web pages, describing user preferences, representing user consumption habits, and defining other information (e.g. demographics models) for targeting a specific audience. This also allows describing segmented content. Segmentation Metadata is used to edit content for partial recording and non-linear viewing. In this case, metadata is used to navigate within a piece of segmented content.

4.3.3. Broadcast content metadata description (Dublin Core, SMTPE)

Dublin Core.

The Dublin Core Metadata Element Set represents a simple resource description record. It is intended to provide a foundation for electronic bibliographic descriptions to improve structured access to information on the Internet. It aims to facilitate the description, organization, discovery, and access of network information resources. The fifteen elements are TITLE, CREATOR, SUBJECT, DESCRIPTION, PUBLISHER, CONTRIBUTOR, DATE, TYPE, FORMAT, IDENTIFIER, SOURCE, LANGUAGE, RELATION, COVERAGE and RIGHTS.

Eight of the 15 elements that make up the Dublin Core metadata set can be "refined" by the addition of one or more qualifiers (e.g. title.alternative) while sets of permitted encoding schemes (defined using the option scheme attribute) have been identified for 10 of the elements.

SMPTE Metadata Dictionary by SMPTE (Society of Motion Picture and Television Engineers) [SMTPE].

The SMPTE Metadata Dictionary acts as dictionary of so-called 'audiovisual descriptors' for the production environment. The Dictionary covers the whole audiovisual production process: pre-production, post production, acquisition, distribution, broadcasting, storage and archiving of digital audiovisual material.

The Dictionary has been designed to allow flexibility in capturing metadata and exchanging it between several applications using a standardized hierarchy of Universal Labels, which are grouped in classes. Metadata Classes are defined as a collection of metadata elements with common characteristics or attributes. The dictionary also includes Additional Classes for user defined metadata. Additionally, The Dictionary contains information on the required format of metadata values and the allowable range of values.

4.3.4. News metadata description (NewsML, SportsML, ProgramGuideML)

NewsML is an XML based, media independent, structural framework for news. It is capable of representing news in all the various stages of its lifecycle in an electronic service environment - e.g. in and between editorial systems;

between wire service providers and media clients;

between original publishers and aggregators / syndicators;

between news service providers and ultimate consumers of news.

NewsML is intended for use in electronic news production, archiving and delivery and as such does not specifically set out to meet the needs of paper-based news publishing. It is intended that NewsML is able to include features required for paper-based publishing and other specific production environments by including external definitions designed for this purpose.

NewsML is not necessarily intended as a format for editing or creation of news, though it is intended as the basis for such formats. It is recognised that the specific demands of individual organizations and production systems will require proprietary extensions to the base NewsML to be effective in this role.

SportsML aims to be the global XML standard for the interchange of sports data. Designed to be as easy to understand and implement as possible, SportsML allows for the exchange of sports scores, schedules, standings, and statistics for a wide variety of competitions.

ProgramGuideML aims to be the global XML standard for the interchange of Radio/TV Program Information based on NewsML. Designed to be as easy to understand and implement as possible, ProgramGuideML allows all for the exchange of Radio/TV information for news publishers and broadcast stations -- program tables(listings), pictures, commentaries, broadcast news, and normative program information.

4.3.5. Metadata exchange (P/Meta, MXF, AAF)

P/Meta by EBU (European Broadcast Union) [PMeta].
P/META is a metadata standard being developed for professional media organizations. It is aiming to build a data model for the exchange of program material between various European broadcasters; and also plans to design a standard approach to structuring information related to media items or objects and to their exchange between process stages and business entities.

MXF (Metadata Exchange Format) [MXF].
The Material eXchange Format (MXF) is an open file format targeted at the interchange of audio-visual material with associated data and metadata. It has been designed and implemented with the aim of improving file based interoperability between servers, workstations and other content creation devices. These improvements should result in improved workflows and result in more efficient working than is possible with today's mixed and proprietary file formats.

AAF (The Advanced Authoring Format) [AAF].
AAF is file format that permits the exchange of essence (picture, sound, video or any other forms) and metadata between multimedia authoring tools. Its major target is to be used as exchange format between different vendors' TV postproduction NLE systems. It is combined of three main parts; AAF object specification, AAF Low level container specification and AAF software development Kit (SDK) reference implementation. The AAF object specification defines logical contents of objects and objects' relations. AAF Low level container specification describes how each object is stored on the disk. AAF SDK Reference implementation is programming tool that lets client applications to access data stored in an AAF file.

4.3.6. Content description framework (RDF)

Resource Description Framework (RDF) by W3C.
The Resource Description Framework (RDF) regroups a wide range of applications from library catalogues and worldwide directories to syndication and aggregation of news, software, and content to personal collections of music, photos, and events and uses the XML language as an interchange syntax.

4.3.7. e-learning content metadata description (SCORM, LOM)

Sharable Content Object Reference Model (SCORM) [SCORM1.2].

The Shareable Content Object Reference Model Initiative (SCORM) is a set of XML based specifications which has been developed to support learning technologies. SCORM aims to provide an integrated suite of e-learning capabilities, which allow the interoperability, accessibility and reusability of Web-based learning content.

SCORM consists of three main elements:

- an Extensible Markup Language (XML)-based specification to represent course structures.
- a set of specifications relating to the run-time environment, including an API and content to Learning Management System (LMS) data model.
- a content launch specification and a specification for the creation of meta-data records for courses, content, and raw media elements.

Learning Object Metadata (LOM) by Learning Technology Standards Committee of the IEEE [LOM].

- LOM aims to specify the syntax and semantics of Learning Object Metadata, defined as the attributes required to fully/adequately describe a Learning Object.
- Learning Objects are defined here as any entity, digital or non-digital, which can be used, re-used or referenced during technology supported learning.
- Examples of technology supported learning include computer-based training systems, interactive learning environments, intelligent computer-aided instruction systems, distance learning systems, and collaborative learning environments.
- Examples of Learning Objects include multimedia content,instructional content, learning objectives, instructional software and software tools, and persons, organizations, or events referenced during technology supported learning.
- The Learning Object Metadata standards will focus on the minimal set of attributes needed to allow these Learning Objects to be managed, located, and evaluated.
- The standards will accommodate the ability for locally extending the basic fields and entity types, and the fields can have a status of obligatory (must be present) or optional (maybe absent).
- Relevant attributes of Learning Objects to be described include type of object, author, owner, terms of distribution, and format.
- Where applicable, Learning Object Metadata may also include pedagogical attributes such as; teaching or interaction style, grade level, mastery level, and prerequisites.
- It is possible for any given Learning Object to have more than one set of Learning Object Metadata.
- The standard will support security, privacy, commerce, and evaluation, but only to the extent that metadata fields will be provided for specifying descriptive tokens related to these areas; the standard will NOT concern itself with how these features are implemented.

4.4. Conclusion

The state-of-the-art in descriptions for adaptation has been presented in this section. MPEG-21 Part 7 Digital Item Adaptation (DIA), tools in MPEG-7 Part 5 Multimedia Description Scheme (MDS) relevant for adaptation, and other relevant standards to UMA are introduced with a large number of usage examples. They cover a wide range of aspects that influence the multimedia content adaptation process including usage environment description, description for transcoding and QoS and description for personalization. This section could play a role to give a hint to identify the necessary descriptions for UMA application designers.

5. Universal Multimedia Access System Designing

This section analyzes what should be considered in Universal Multimedia Access systems designing. In UMA system designing, it is necessary to decide how to locate the adaptation engines, contents and descriptions considering the available computational resources. On the other hand, UMA systems use more and more personal information to provide better services. It is essential in designing UMA systems to consider a mechanism to protect that information. Architectural issues are presented in 5.1 and potential problems on privacy in UMA systems are given in 5.2.

5.1. Designing how to locate engines, contents and descriptions in UMA systems

5.1.1. What is UMA system designing?

UMA system designing can be defined as a problem of how to locate the engines, contents and descriptions that are necessary in UMA systems. The main issues to be considered to solve this problem include the (computational) resource constraints, effectiveness of adaptation process and the target application. The engines, contents and descriptions can be located on the server side, client side, intermediate server (proxy server) or combination of them.

UMA system designing problem can be given as follows;

<u>What to locate?</u>

1. *Adaptation engines.*
2. *Multimedia contents.*
3. *Content descriptions.*
4. *Usage environment descriptions.*
5. *Provider environment descriptions.*

<u>Where to locate?</u>

a. *Server side.*
b. *Client side.*
c. *Intermediate server (proxy server)*
d. *Combination of a. - c.*

5.1.2. Classification of adaptation system architectures

It is quite important to define where to locate the adaptation engine as the adaptation engine requires the most computational power. The adaptation engine can be located in the server, the client, intermediate server (proxy server) or combination of them. Each of them has its advantages and the architecture has to be determined considering the available resources and the target application. Here we discuss the three possibilities in where to adapt the content; server-side, client-side or proxy-based adaptation.

5.1.2.1. Adaptation on server-side

In server-side adaptation, the server analyzes all the environment descriptions and adapts the contents in an optimal way under given constraints. The adaptation could be performed on server-side in case the client terminal has capability limitations (e.g. processing capabilities, access methods, display capabilities) and it is not possible to receive or adapt the content source. The merits and defects in server-side adaptation are given in the following;

Merits:

- Enables flexible decision of adaptation strategies (use of transformation engine, variation selection engine, content selection engine, etc,) depending on the environments.
- The server can control how the contents can be consumed.

Defects:

- There are cases where too much load is put on the server.
- Requires a lot of computational resources on the server.

5.1.2.2. Adaptation on client-side

In client-side adaptation, the client receives the requested contents and adapts to its capabilities and constraints by itself. Client-side adaptation is useful in where the contents are distributed in a wide range and every consumer receives the same content (e.g. broadcast content receiver). It is also useful where there is no server for content delivery and several clients exchange their contents (e.g. peer-to-peer service). The merits and defects client-side adaptation are given in the following;

Merits:

- Usage environment descriptions are not necessary to be sent to the server. Reduces necessary network resources.
- Low cost implementation of the server.

Defects:

- A lot of computational power is required in the client devices. This makes the device manufacturing cost expensive.
- The server cannot know the how each user interacts with the content.

5.1.2.3. Proxy-based adaptation

In proxy-based adaptation, the proxy server receives all the environment descriptions and adapts the contents in an optimal way under given constraints. The proxy server requests the server the necessary contents, adapts them and send them to the client. The merits and defects in proxy-based adaptation are given in the following;

Merits:

- There is no need to change existing clients and servers.
- Enables distribution of computational loads. Separation of content management and adaptation.

Defects:

- More negotiations among servers, proxies and clients are required.
- The system structure becomes more complex.

5.1.3. Storage location of usage environment descriptions

It is also quite important to determine where to locate the usage environment descriptions as these descriptions contains a lot of personal information. The usage environment description can be located in the server, in the client, in intermediate servers, or combination of them. Each of them has its advantages and the location has to be determined considering the privacy issues besides the available resources and the target application. Privacy issues are described in detail in 5.2. Here we discuss the three possibilities in where to locate the usage environment descriptions; server side, client side or intermediate server.

5.1.3.1. Storage in server side

To store the usage environment descriptions on the server side raises important privacy concerns. On the other side, better UMA services can be provided by having these descriptions on the server side. The main problem remains on how to reconcile personalization and privacy.

5.1.3.2. Storage in client side

To store the usage environment descriptions on the client side and not send any of them outside could be the best way to keep privacy. However, this limits extremely the possibilities on providing good UMA services. This could be a good solution if the adaptation is processed on client side. If the adaptation is on server or proxy side, then those descriptions have to be sent there. This means that secure transmission technologies like encryption or anonymized transmission would be necessary.

5.1.3.3. Storage in intermediate server

The usage environment descriptions could also be stored in an intermediate server. This server could be a trusted third party, a server that anonymise the descriptions, or some trusted communities or networks (e.g. social networks).

5.2. Privacy in Universal Multimedia Access systems

This section analyzes potential privacy problems in Universal Multimedia Access systems. It is obvious that quite a lot of personal information (e.g. user preference, usage history, access information, location information, user's terminal) is necessary for content adaptation. For better context aware multimedia content delivery and access service, more personal information is necessary. At the same time, this means that privacy concerns would become aware in UMA services. UMA will never become a practical service without considering privacy. It is essential for Universal Multimedia Access systems to have a privacy protection mechanism.

Currently, the main activities in MPEG-21 target two major topics. The first one is Intellectual Property Management and Protection (IPMP), which includes digital rights management (DRM) matters, and the other one is Universal Multimedia Access (UMA), about a seamless access to multimedia contents from anywhere at anytime. However, one important topic is still missing in MPEG-21. Privacy.

In this section we try to identify potential problems on privacy in Universal Multimedia Access systems. Activities and standards relevant to privacy are also introduced to make clear what should be considered when designing a UMA system with privacy protection mechanism.

5.2.1. What is privacy?

Privacy protection is an emerging issue in not only in UMA services, but also in most of ubiquitous services. Privacy protection is quite important for service providers to be kept trusted and for consumers to be sure that their privacy is protected while using the service.

Before analyzing how privacy should be protected in UMA system, we need to define what privacy is. There are so many different definitions of privacy depending on its meaning within a context. Some popular definitions are "The right to be left alone." (1890. Louis Brandeis), and "a) The state of being in retirement from the company or observation of others. b) Freedom from unauthorized intrusion." (Webster's dictionary).

Roger Clarke 0 categorized privacy into four interests:

- **Privacy of personal data**: Individuals claim that data about themselves should not be automatically available to other individuals and organizations, and that, even where data is possessed by another party, the individual must be able to exercise a substantial degree of control over that data and its use [data privacy, information privacy].

- **Privacy of personal communications:** Individuals claim an interest in being able to communicate among themselves, using various media, without routine monitoring of their communications by other persons or organizations [interception privacy].

- **Privacy of the person:** Integrity of the individual's body. eg) Compulsory immunization, blood transfusion without consent, compulsory provision of samples of body fluids, compulsory sterilization, etc. [bodily privacy].

- **Privacy of personal behavior:** Sexual preferences and habits, political activities, religious practices [media privacy].

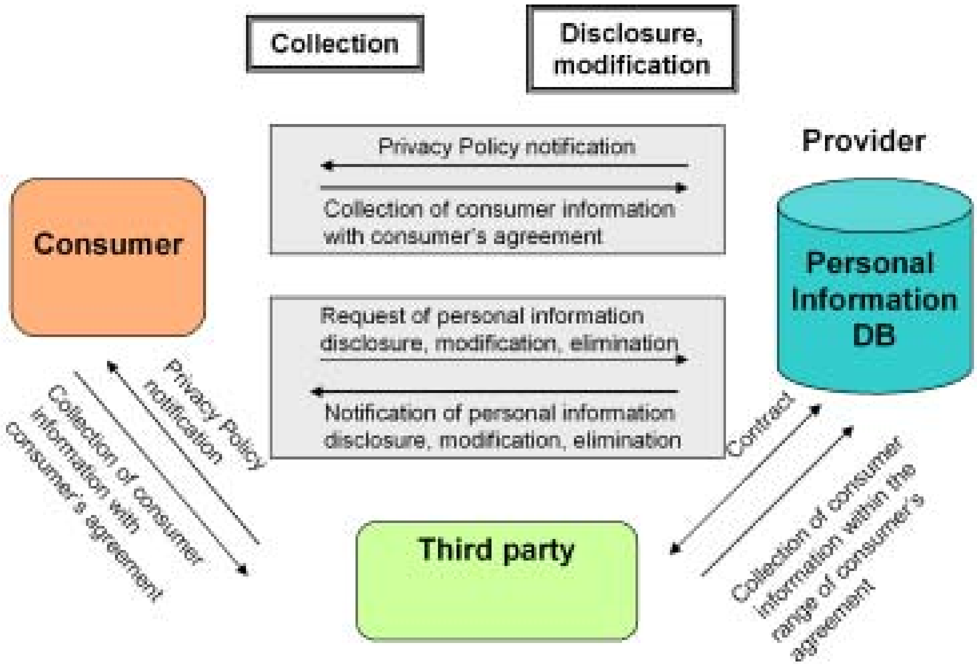


Figure 33. Framework of personal information protection

The term **'information privacy'** refers to the combination of communications privacy and data privacy 0. Information privacy refers to the claims of individuals that data about themselves should generally not be available to other individuals and organizations, and that, where data is possessed by another party, the individual must be able to exercise a substantial degree of control over that data and its use 0. Privacy protection could be defined as a process of finding appropriate balances between privacy and multiple competing interests 0.

In this report, we focus on "information privacy" as this is exactly what UMA service users would be concerned.

5.2.2. Privacy protection framework

Figure 33 illustrates a typical framework of how to handle personal information for privacy protection. This framework is based on two principles [Koizumi02].

1. Notification of usage purpose and user's agreement are essential (Informed consent).
2. Requests on personal information disclosure, modification and elimination from the person "should" be followed (not "must" because there are exceptions, e.g. data that government controls.).

Considering the interface between provider and consumer, in case the provider directly contacts with the consumer, the provider notifies his/her privacy policy (policies of the provider on how the personal information is processed and used) to the consumer and collects consumer information after his/her agreement. The provider should follow the requests on personal information disclosure, modification and elimination from the consumer. In case the provider outsource to a third party the work of contacting with consumers, this third party notifies privacy policy to the consumer and collects consumer information with consumer's agreement. Then, the provider collects consumer information within the range of consumer's agreement with the third party. Management of personal information in provider side must be kept correct and secure. Transfer of personal data to a third party is only allowed if the consumer has agreed with the transfer, and only within the range that consumer has agreed.

There are many problems and concerns in this framework.

Provider side:

- How to let the user know our privacy policies.
- How to enforce personal data protection internally.
- What can we do to increase user awareness of what the environments are doing, and how user information is collected and used?
- How can we empower users to manage the ways in which they are represented in the environments, or to limit their exposure when needed?
- How to let the user be aware who knows what about him at what time.
- Too much security issues to provide a service within an acceptable delay.

Consumer side:

- How can I control my personal information?
- How can I know if the provider keeps appropriately personal information?
- Too hard to read and understand all the policies. How can I judge if I can trust the provider or not? (Many users even don't read them.)
- How can I access with anonymous name?
- How can I know how data is being used? Who has access to personal information? When did they access it? From where? What did they look at? How often do they view this information?
- How can I know when a privacy violation occurs?
- How can I know if they are keeping their privacy policies?

It is obvious that a check system which ensures that the provider correctly implements the policies is necessary. At the same time, some transparent way of describing and exchanging privacy policies are also necessary. Another problem is that consumers are not so much aware on privacy issues in the virtual space such as internet as on privacy in general.

Many approaches have been tried to solve these problems and keep better privacy. Some of them are technical, some are environmental.

- **Anonymizers**: Anonymizers allows the consumer to browse the Internet using an intermediary to prevent unauthorized parties from gathering your personal information. On the other hand, it makes difficult to use personalized service.

- **Platform for Privacy Preferences (P3P)**: A standard that enables to express privacy practices in a standard format that can be retrieved automatically and interpreted easily by user agents (see 0).

- **Privacy policies**: Privacy policies are created to inform users of a site's data collection, use and disclosure practices. Posting privacy policies is essential in building trust between Web sites and their users.

- **Privacy interest groups**: Several groups have been founded to raise public awareness on privacy by providing information about privacy.

5.2.3. Privacy in UMA systems

Basically, UMA systems are based on user's access to the contents. If we assume the usage of MPEG-21 DIA descriptors, at least, the following information can be obtained from user's access;

- Accessed location, time, viewed content information.
- User's terminal information.
- Usage history of contents.
- User preference data itself or preference data by analyzing accessed content history and location information.)
- Daily lifestyle by analyzing location and time information.
- Behavior analysis from access log.
- MPEG-21 DIA descriptions itself, especially usage environment descriptions.

Those data, especially user preference and usage history, are very useful for personalized services but contain a lot of personal information at the same time. We should protect in UMA systems any personal information of the user including all the information above and information obtained by processing that information. However, that information is essential for good UMA service and we cannot just get rid of them. We need to balance between privacy and UMA service quality.

5.2.4. Requirements in UMA system designing

In UMA systems, MPEG-21 DIA description, access information and information derived by analyzing them are the information which needs to be protected.

We should consider the following things when designing a UMA system;

1) Informational self-determination framework.
   How to enable user to control their user information (modify, access, delete).
2) Informed consent to user.
   (e.g. Privacy policy).
3) Transparent privacy policy exchange.
   As well as capability configuration, policy exchange is necessary. One way could be MPEG-21 DIA and P3P harmonization. Some work in implementing CC/PP and P3P have been done 000, which would be a good reference for MPEG-21 DIA and P3P harmonization.
4) How to allocate database and servers.

e.g. Put all information into user terminal?
Store personal data with anonymous name?

5) Enforcement of **internal** personal data protection.

e.g. How to guarantee that the service provider doesn't distribute personal information.
How to reduce damage in case partial personal information are released accidentally or on purpose.

5.2.5. Standards relevant to privacy

TRUSTe

TRUSTe is a non-profit organization dedicated to enabling individuals and organizations to establish trusting relationships based on respect for personal identity and information. TRUSTe gives a "trustmark" to sites that adhere to established privacy principles (Figure 34).
The technology's guiding principles are as follows;

- **Adoption and implementation of a privacy policy** that takes into account consumer anxiety over sharing personal information online.
- **Notice and disclosure** of information collection and use practices.
- **Choice and consent**, giving users the opportunity to exercise control over their information.
- **Data security and quality and access** measures to help protect the security and accuracy of personally identifiable information.

Figure 34. trustmark of TRUSTe

Some other privacy seal organizations are;
• PrivacySecure www.privacysecure.com
• BBBOnline Privacy www.bbbonline.org
• Privacy Rights Clearinghouse www.privacyrights.org

P3P (Platform for Privacy Preferences 1.0)

The Platform for Privacy Preferences Project (P3P) **0** enables Web sites to express their privacy practices in a standard format that can be retrieved automatically and interpreted easily by user agents. P3P policies consist of statements made using the P3P vocabulary for expressing privacy practices. P3P policies also reference elements of the P3P base data schema -- a standard set of data elements that all P3P user agents should be aware of.
The P3P specification defines:

- A standard schema for data a Web site may wish to collect, known as the 'P3P base data schema'
- A standard set of uses, recipients, data categories, and other privacy disclosures

- An XML format for expressing a privacy policy
- A means of associating privacy policies with Web pages or sites
- A mechanism for transporting P3P policy statements over HTTP.

The P3P specification includes a mechanism for defining new data elements and data sets, and a simple mechanism that allows for extensions to the P3P vocabulary.

Basic P3P interaction is as follows;

1. The agent requests a Web page from a service.
2. The service responds by sending a reference to a P3P **policy-reference** in the header of its HTTP response. A policy-reference file lists parts of a Web site and the URIs of their corresponding privacy policies. A policy consists of one or more statements about a service's privacy practices.
3. The agent fetches the policy-reference file and determines the URI of the policy that applies to the requested page.
4. The agent fetches the policy, evaluates it according to the user's **ruleset** (which represents her **preferences**) and determines what action to take (e.g., simply informing the user about the privacy policy in place, or prompting her for a decision).
5. In some implementations, a match between the user's preferences and a site's policy might authorize electronic wallets and other data repositories to (semi-) automatically release information to the service.

APPEL (A Privacy Preference Exchange Language)

APPEL 1.0 **0** specifies a language for describing collections of preferences regarding P3P policies between P3P agents. Using this language, a user can express her preferences in a set of preference-rules (called a **ruleset**), which can then be used by her user agent to make automated or semi-automated decisions regarding the acceptability of machine-readable privacy policies from P3P enabled Web sites.

MPEG-21 Part 5 and MPEG-21 Part 6

MPEG sees a **Rights Data Dictionary 0** as a dictionary of key terms which are required to describe rights of all Users, including intellectual property rights, that can be unambiguously expressed using a standard syntactic convention, and which can be applied across all domains in which rights need to be expressed.

A **Rights Expression Language 0** is seen as a machine-readable language that can declare rights and permissions using the terms as defined in the Rights Data Dictionary.

The Rights Expression Language is also intended to provide a flexible interoperable mechanism to ensure personal data is processed in accordance with individual rights and to meet the requirement for Users to be able to express their rights and interests in a way that addresses issues of privacy and use of personal data.

6. Challenges in Universal Multimedia Access

This section summarizes the challenges in Universal Multimedia Access. The most important challenges in UMA systems are in the following; User-centric multimedia content adaptation, evaluation metrics, and system designing considering the target application, necessary descriptions, privacy protection, digital rights management and content usage control. It is also a big challenge for the success of UMA to promote the use of standardized metadata (e.g. MPEG-7/MPEG-21) for those who are potential UMA customers.

6.1. User-centric multimedia content adaptation - how can the best experience be provided? -

The most important thing in UMA services is to adapt and deliver the contents in a way that maximizes the user satisfaction. The new challenges exist in combining multiple engines, contents and variations in order to provide the best experience to the user. On the content side, new problems arise when more than one content or variation or modality is necessary to be presented. The main new problem is how to jointly adapt multiple contents and variations to maximize the experience of the user within the given restrictions. More details on optimization across multiple adaptation engines are presented in 6.1.1.. 6.1.2 describes on joint adaptation of multiple contents and variations.

6.1.1. Optimization across multiple adaptation engines

As described in section 3, each of the transformation engines, variation selection engines and content selection engines has its own challenges and there is still a lot to be done. In addition to these challenges on each single engine, combination of multiple engines also needs to be considered to provide the best experience to the user. When multiple engines are required to be combined, new challenges arise. The new challenges exist in how to combine multiple engines in an optimal way.

Figure 35 compares an example of conventional adaptation process and using multiple engines. The conventional approach is to perform each engine one by one in a cascade way. Figure 35 (i) shows an example of an approach for adaptation using multiple engines presented by Steiger [Steiger03]. This approach first selects the best contents based on user profiles, then selects the best variation based on the terminal and network capabilities, and finally transforms the selected content to fit the display capability. However, the best selection of personalized contents would depend not only on user profiles, but also considering the terminal and network capabilities in order to determine how many contents can be selected. There would also be a tradeoff between the computational cost of the whole adaptation procedure and the quality of the adapted contents. The new challenge is how to determine the strategy of optimal adaptation and then perform the adaptation based on the determined strategy. The strategy determination includes how to combine any transformation engines, any variation selection engines and any content selection engines to optimize the overall adaptation quality and performance considering all the given parameters (Figure 35 (ii)).

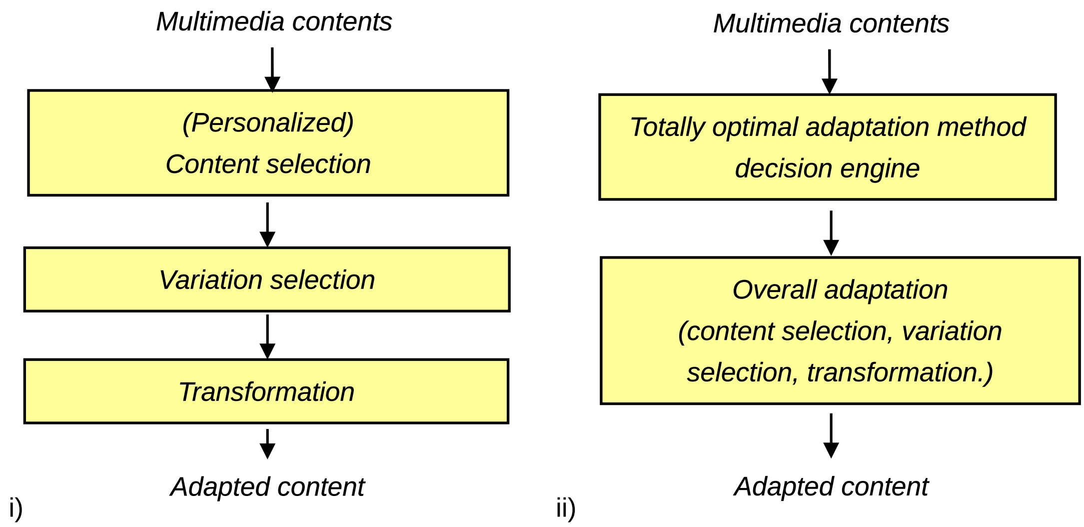


Figure 35. Example of a conventional method and challenges in multimedia content adaptation using multiple engines.

6.1.2. Joint adaptation of multiple contents and variations

On the content side, when more than one content or variation or modality is necessary to be presented, new problems also arise. The main new problem is how to jointly adapt multiple contents and variations to maximize the experience of the user within the given restrictions.

6.1.2.1. Joint adaptation of multiple contents

Joint adaptation is necessary when more than one content are requested and have to be delivered to the consumer. Besides the optimization of adapting each content, the optimization of all the adaptation process and presentation of multiple contents has to be considered. For instance, how to deliver and present all the recommended TV programs considering the user environment. Some works are presented in multi-channel delivery of broadcast contents adapting a fixed bandwidth. This is a simple example of multiple bitstream transcoding in signal domain. Joint transcoding of multiple contents using other engines like transmoding, resolution and length transformation, etc. is also necessary.

6.1.2.2. Joint adaptation of multiple variations

Joint adaptation is also necessary when more that one variation or modality of a single content has to be delivered to the consumer. The optimization of all the adaptation process and presentation of multiple variations has to be considered. This adaptation technology is useful to browse a content with several modalities. For example browsing a music clip with its lyrics in PC, the optimal adaptation and presentation of “audio + video + artist info + lyrics ” is necessary. Even in mobile devices, a joint adaptation and presentation of “audio + lyrics ” would be necessary to control the audio quality and length of lyrics considering the network conditions.

6.1.2.3. Presentation of the adapted multimedia contents

How to present the transformed or selected content(s) to the user to maximize their satisfaction is also another challenge in user-centric adaptation.

The first problem is that it is necessary to analyze what kind of view the user really wants, not only in terms of terminals and networks, and how the contents should be presented to the user. Some factors for visualization and presentation could be how easy to access, to browse, to understand the contents, and how much the presentation way fits to his/her preferences (it quite depends on the application.)

The other problem is that after adaptation based on transformation or selection, there would be cases that too many contents are selected and makes the user difficult to access to the desired content. Structured content visualization and metadata visualization could be a good assistance method for easy access to the desired content.

6.2. Evaluation metrics - how can the quality of experience be evaluated? -

Another big challenge is how to evaluate the quality of experience. There are two factors in evaluation, the evaluation of the adapted content and the determination of adaptation policies.

6.2.1. Metrics to evaluate the quality of the adapted contents

The evaluation metrics of the quality of the obtained adapted contents for the user is emerging. It is obvious that neither PSNR nor compression ratio are appropriate for measuring the quality of experience of the user. The evaluation metrics should include some measurement of user satisfaction, for example, the quality of the context in the image/video, how much the provided contents fit their preferences, and any other factors that effects the user experiences, and so on. The challenges can be divided in two factors, the evaluation of perceptual quality, and the evaluation of semantic quality. The challenges are in the following;

- *How can the perceived quality be evaluated?*
  - How to measure the effect of signal reduction to perception.
  - How to measure the effect of spatial / temporal resolution reduction.
  - How to measure the effect of artifacts (e.g. coding artifacts).

- *How can the user satisfaction on the semantics of the given content be evaluated?*
  - How to measure the quality of the selected modalities.
  - How to measure the quality of the summarized / personalized contents.
  - How to measure the quality of the context.

6.2.2. Metrics to determine how to adapt the contents

The other challenge is how to adapt the content(s) in a way that provides the best experience to the user. The challenges are presented in the following;

- *Definition of rules for adaptation.*
  - How to create generic rules for adaptation.
  - How to consider the tradeoff between complexity and quality.
  - How to balance between spatial and temporal resolution.

- *Definitions of content importance value.*
  - How to define the value of each variations.
  - How to define the value of the adapted contents.
  - How to define the best variation, best modality, etc.

6.3. UMA System designing - How can we make UMA systems practical? -

All of the aspects described in the following should be considered for UMA systems. It is quite important to consider all of them in the beginning of a UMA system designing process instead of developing each of them separately and combining them all afterwards.

- How to locate of engines, descriptions and contents.
- Available computational resources and restrictions.
- Identification and selection of necessary descriptions.
- Privacy protection
- Rights management.
- Content usage control.
- Target application.

6.3.1. Efficient creation of content metadata and descriptors for adaptation

For efficient creation of content metadata and descriptors for adaptation, two things are important. To identify which ones are necessary and how to prepare them efficiently.

Before adaptation, the contents and descriptions need to be prepared. A lot of topics should be considered in preparation of descriptions and variations in an efficient way. Some example challenges are given in the following;

- What kinds of descriptions and variations are necessary for adaptation, for real-time transformation, for personalization, for transcoding, for a specific application?
- How to give an adequate keyword to describe the contents?
- How to manage variations from a single content?

MPEG-7 and MPEG-21 provide rich tools to describe contents, variations and environments. As the requirements for the system depends quite a lot on the application, it is essential to identify the necessary descriptions and variations for each application. After identifying what kind of descriptions and variations are necessary, it is very important to create them in an efficient way. Some example challenges are given in the following;

- Which are the metadata that can be created automatically, manually, semi-automatic?
- How can they be created automatically?
- How to generate variations within a reasonable cost?
- How to update variations when there are changes in the source content?

6.3.2. Privacy protection, digital rights management and content usage control

Privacy protection is essential to make UMA systems practical. As discussed in section 5, the main problems on privacy in UMA exist in how to reconcile personalization and privacy; the challenges in privacy are as follows;

- *Informational self-determination framework.* (Everyone has the right to know who is knowing what about him at what time.)
    - How to enable user to control their user information (modify, access, delete).
    - Informed consent to user. (e.g. privacy policy).
    - Transparent privacy policy exchange.

- *How / where to allocate usage environment descriptions.*
    - All information in client terminal?
    - Third party? Anonymizer? Trusted party to keep the information?
    - Restricted use in communities or social networks?

- *Enforcement of internal personal data protection*
    - e.g. How to guarantee that the service provider doesn’t distribute personal information.

It is important to design a system in a way that increases user awareness of what the system environments are doing, and how user information is collected and used. It is also important to empower users to manage their personal data.

Another relevant big challenge is how the digital rights management and the way for consuming the contents can be controlled. Example challenges are given in the following;

- Quality restrictions.
    - *Commercial film not allowed to be sent with low resolution.*

- Restrictions on the length and context to be sent.
    - *e.g. some (part of) the content cannot be sent to some people of a certain category. (*Not to deliver a climax scene as a movie preview…)

- Definitions of the rights holders for derived variations.
    - *e.g. the transcoded or transmoded variation belongs to whom? Summary of the contents belongs to whom?*

6.3.3. UMA Application

To find a killer application is another very big challenge in UMA. There are two obvious application domains in UMA; streaming applications and universal access applications.

A. Streaming applications (streaming audiovisual media resources):

Streaming refers to the ability of an application to play synchronized media streams like audio and video streams in a continuous way while those streams are being transmitted to the client over a data network. Applications of streaming services can be classified into on-demand and live information delivery applications. Examples of the first category are music and news-on-demand applications. Live delivery of radio and television programs are examples of the second category [PSS234]. It is obvious that the contents are preferable to be delivered in a way which maximizes the user satisfaction.

B. Universal access applications (seamless access to media resources):

Universal access refers to the ability of an application that allows access to multimedia contents over any type of network with any device from anywhere and anytime (universally). It is also obvious that the contents are preferable to be accessed in a way which maximizes the user satisfaction.

For more specific UMA applications, it is important to make clear "what" the application delivers to "whom".

7. Conclusion

This part analyzed the state-of-the-art technologies in Universal Multimedia Access (UMA), and tries to identify the key issues of UMA and challenges that still remain to be resolved in UMA.

There are three key issues in UMA; User-centric multimedia content adaptation, (standardized) description necessary for adaptation, and system designing considering the target application, necessary descriptions, privacy protection, digital rights management and content usage control.

We have categorized the adaptation engines for user-centric multimedia content adaptation into four types considering their functionality. They consist of adaptation of the content by transforming it on the fly, by selecting the content variation, by selecting the preferred content, and by combining some of them, all in a way that the best possible experience is provided to the consumer. State-of-the-art technologies and their problems for each of them were presented. To achieve user-centric multimedia adaptation, there still remain a lot of unsolved problems which includes how to optimize the use of multiple adaptation engines in a way that provides the best experience to the user, how to jointly adapt multiple contents and variations, how to present the transformed or selected content(s) to the user to maximize their satisfaction and how to measure the “quality of experience”.

The state-of-the-art in description necessary for adaptation, which includes MPEG-21 Part 7 Digital Item Adaptation (DIA), some tools in MPEG-7 Part 5 Multimedia Description Scheme (MDS), and other relevant standards, have been reported with a large number of usage examples. They cover a wide range of aspects that influence the multimedia content adaptation process including usage environment description, description for transcoding and QoS and description for personalization.

As a large amount of personal information is required for content adaptation, privacy protection must also be considered in UMA systems. Activities and standards relevant to privacy were analyzed to make clear what should be considered when designing a UMA system with privacy protection mechanism. The key problems on privacy in UMA systems are as follows; informational self-determination framework, informed consent to user, transparent privacy policy exchange, the allocation of databases and servers, and the enforcement of internal personal data protection.

Challenges in UMA include finding a good target UMA application. It is also important to consider all the aspects in UMA, which include adaptation engines, necessary descriptions and privacy protection, in the beginning of a UMA system designing process instead of developing each of them separately and combining them all afterwards. Key problems in system designing in UMA are to find a good UMA application, to identify necessary descriptors, and to locate databases and servers considering the complexity, functionality and privacy protection.

Part II: Joint Adaptation of an Audiovisual Content and its Metadata

8. Introduction

The recent innovation of information technologies has enabled us more and more opportunities to access and consume audiovisual contents. Thanks to the advancement of computational power and storage capability, the number of new appliances to consume these audiovisual contents like digital TV, PC, PDA, mobile phone, portable devices, and so on, is increasing tremendously. At the same time, more and more audiovisual contents are distributed with some metadata. Metadata helps not only managing and searching contents. The metadata also helps the consumer browsing audiovisual contents by providing additional information of the content, and also enables flexible content navigation by linking each part in the metadata to the corresponding part of the content. For example, many news programs present their explanations in text in addition to the news video clip. Movie trailers are presented on the web with their detailed information on the same page. Thus, the way of consuming audiovisual contents has changed dramatically because of the richness of the created contents and the variety of new consuming devices. As there is a wide variation of the capabilities of these devices like screen size, CPU power, playable formats, display capabilities (color, fonts, etc.), technologies to bridge the gap between the rich contents and the capabilities of the new devices are emerging.

Since screen size is one of the most relevant limitations on consuming devices, it is important to maximize the user experience in presenting an audiovisual content with metadata in screens having any size. It is obvious that one presentation doesn't fit all. The current and typical approach to this problem, called "variation selection", is to prepare different versions of the content for each type of devices. The available choices are presented to the user and the user selects the best one, or the system automatically determines the user's terminal and leads him/her to the most appropriate variation [Mohan99]. The problem is that too much task is necessary to prepare and manage different variations for all existing devices. It is especially hard to update all the derived variations in case of modifying the original source. There is a strong demand for adapting these contents automatically to each device with any screen size.

This paper proposes a framework for joint adaptation of audiovisual contents and its metadata. The presentation of the audiovisual content and its metadata are balanced to fit the given screen size in a way that maximizes the browsing experience. This paper is organized as follows. Section 2 gives the related works. The overview of the proposed framework is introduced in section 3. Section 4 explains how to jointly adapt multimedia contents including different modalities and balance their presentation. The adaptation method for each modality is described in Section 5. Section 6 presents a prototype system using this framework and evaluation results. Conclusions are given in section 7.

9. Related work

Some efforts have been made on multimedia content adaptation like image adaptation, video adaptation and web page visualization adaptation.

For image adaptation, the most straightforward way is to simply scale the image to fit the given screen size. However, if the screen size is very small, it becomes too difficult to recognize the involved context. Some works are presented to display large images to small screens. The basic approach is to present some important part of the image instead of presenting the whole image. Lee [Lee01] proposed a framework to prioritize the presentation of the region-of-interest (ROI) inside the image. The image is resized fitting the client display size considering the object importance inside the image. Importance values are added beforehand to each block, and the resizing process is performed by combination of cropping and scaling using the proportion that maximizes the content importance values. (same idea as Region-of-Interest(ROI) in JPEG2000.) Liu [Liu03] proposed a rapid serial visual representation (RSVP) approach to browse a large image in a small screen. Instead of just scaling down the image, it is divided in rectangular regions (allowing overlap) including important objects, and they are presented one by one. The optimal order and time of presenting image objects are modeled. Fan [Fan03] prioritizes the regions which catch the visual attention of a human. Face, text, and saliency part is considered to be important and high priorities are assigned on these objects.

For video adaptation, all the methods for image can be extended by considering each frame as an image. In addition to the spatial factor, the temporal factor is also controlled in video adaptation. Mohan [Mohan99] presented a system that controls the modality (to show video or key frames), control the size and color depth for spatial resolution, and the number of keyframes to be displayed considering the constraints. Kim [Kim03] proposed a method for bit-rate adaptation which controls the spatial and temporal resolution. Cavallaro [Cavallaro03] proposed an *object-based conversion method* which controls the coding conditions or decoding order of objects or regions of interest inside the image/video. He also proposed a *description-based conversion*, which delivers/uses just the features extracted from the objects (object identifier and shape information given as an example) when the constraints are very tough.

For multimedia content adaptation, some studies on browsing large web pages into small displays have been presented. Opera browser's "Small-Screen Rendering" [Opera] reformats the page to fit inside the screen width and eliminate the need for horizontal scrolling. The web page structure is analyzed, table columns are extracted, each table is resized to fit the width of the screen and then they are located one by one horizontally from the top in the small browser. Images are rescaled but as all the text information still remains the same, there are difficult cases in browsing the whole page when the page is big. Frayling [Frayling02] and Wobbrock [Wobbrock02] tries to solve this problem by creating a thumbnail of the whole page browsed by PC and present it in small screen. This thumbnail also indicates logical segments and each segment is clickable. The selected segment can be zoomed.

10. Screen size adaptation for multimedia browsing

10.1. Overview

This section presents a framework for browsing video content and its metadata adapting any screen size and identifies which tools are necessary in screen size adaptation. For maximizing a consumer's browsing experience of a video content with its metadata, it is essential to adapt both the video content and its metadata and to balance them in an optimal way to fit any screen size.

Figure 36 illustrates a typical example of how a specific video content are presented with its metadata. The video content is delivered on the left pane of the screen so that the consumer can watch it. The metadata helps the consumer browsing this video in two aspects. The metadata provides additional information of the video content so that the consumer can understand better the content. It also allows a flexible navigation of the video content by linking each metadata to the corresponding part of the video. This helps access to the desired part of the video content.

Figure 37 shows the concept of screen size adaptation of the content (video with its metadata). When this content is delivered to various devices with different screen size, it is obvious that preparing just one version is not enough to adapt to any screen size. For devices with large screen size like PC, more textual information and/or larger size of a video content can be provided considering the consumer's preference. On the other hand, devices with small screen size like PDA or mobile phones, a smaller video content with less number of textual information should be provided nevertheless the consumer will have difficulties in browsing (e.g. forced to scroll the screen all the time). The presentation of the content should be flexible considering the screen size. Both the video content and its metadata have to be adapted to the screen size in an optimal way. Moreover, this adaptation includes not only "video optimization problem" and "metadata optimization problem". When two modalities are combined, new problems also arise. The main new problem is how to jointly adapt these modalities and balance their presentation to maximize the browsing experience of the user within a given screen size.

10.2. System design

Figure 38 shows the overview of the proposed adaptation system. This system enables the consumer to access the desired video content with its metadata in an optimal way by adapting them to the screen size of his/her device and to his/her preferences.

Three types of data are necessary to be prepared beforehand. First, the **video contents** to be delivered are prepared. The system also includes the **content metadata** of the video content, which describes the structure of the video and the text annotation for both each item and the whole content. The **adaptation tools,** which include hint information for guiding the adaptation process, are also prepared for efficient adaptation of the desired contents to the consumer. The preparation process of these data is described in detail in section 10.3.

First, the consumer requests the desired content. At the same time, the information of the consumer, which includes the consumer's **browsing preferences** and the **screen size** of the consumer's device or browser, are sent to the provider to let the provider know the consumer's capabilities and preferences. Details of the descriptions are presented in section 10.4.

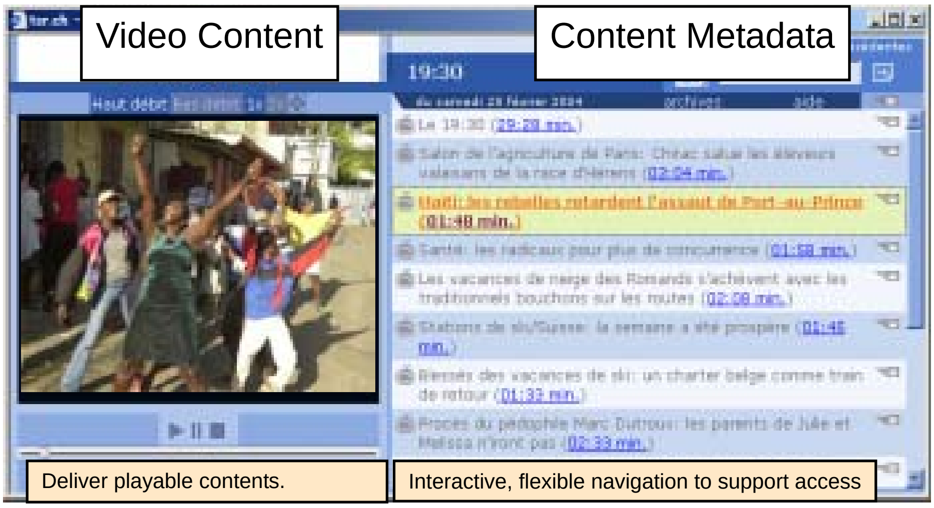


Figure 36. Typical example of video content presentation with its metadata (TSR TV news)

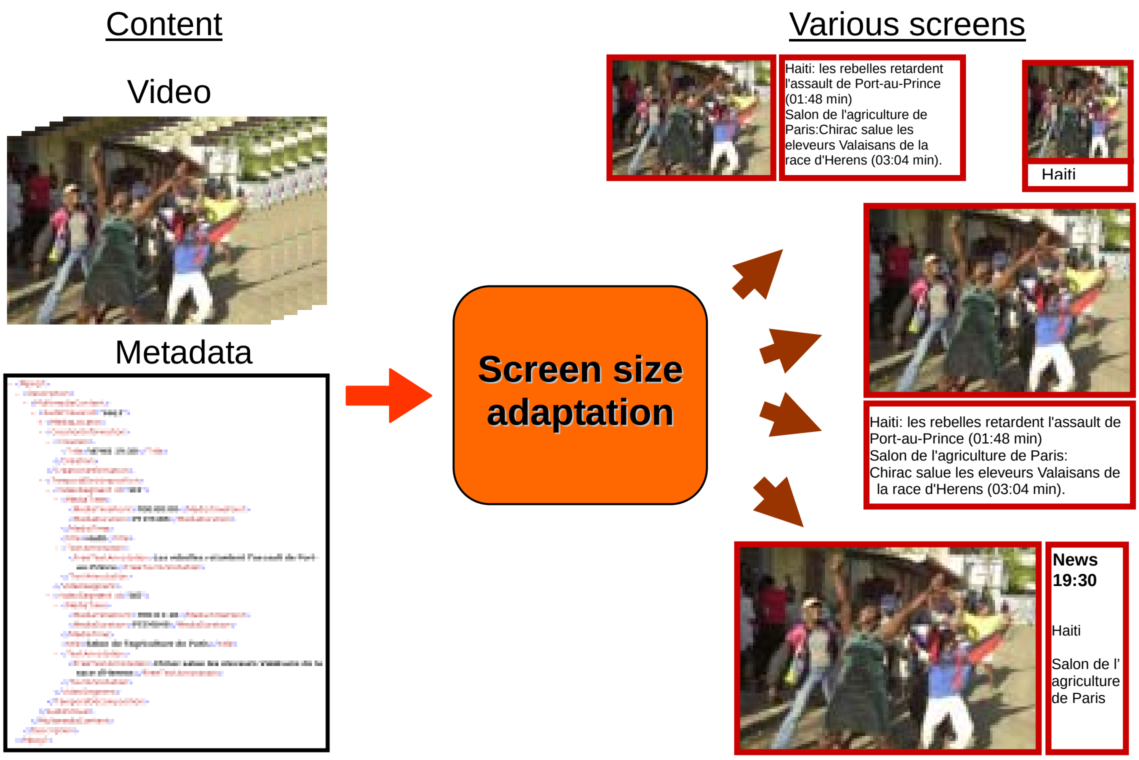


Figure 37. Concept of screen size adaptation

The **adaptation engine** tries to adapt these different modalities and balance their presentation to maximize the browsing experience of the user within a given screen size. More information of this adaptation engine is described in section 10.5. The adapted contents (video + metadata) are wrapped and delivered to the consumer in a suitable description format like SMIL (Synchronized Multimedia Integration Language) or HTML (Hypertext Markup Language) considering their presentation balance. The consumer device receives and visualizes the delivered contents in a way that maximizes the user's experience.

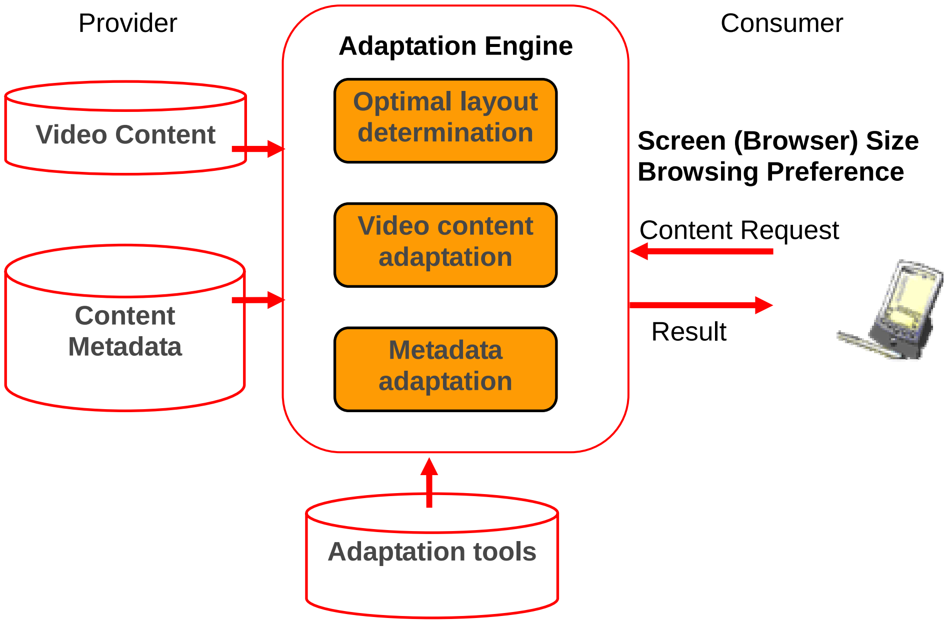


Figure 38. Overview of the adaptation procedure of a video content with its metadata

In system designing, it is important to determine how to locate the adaptation engine, contents and descriptions. There are three possibilities in where to locate the adaptation engine.

1. **Server side:** The server receives the client requests, decides the adaptation process, adapts the contents and sends the adapted contents to the client. The client terminal just has to receive the adapted contents. Good solution for low-cost client terminal and low-cost adaptation.
2. **Client side:** The client terminal receives the contents as is, and the terminal itself adapts the contents to fit its capabilities. Good solution for peer-to-peer service and also for digital broadcast receivers where flexibility of the client is important and also adaptation for each terminal is too expensive.
3. **Intermediate server (proxy server):** The content server receives the request and let intermediate server(s) perform the adaptation decision and the adaptation process. Good for distributed services, for separation of content management and adaptation.

The strategy for system designing has to be selected considering the computational resource constraints and the target application.

It should be noted that to maximize user experience, it is important not only to present the content and its metadata in an optimized way but also to allow users flexible browsing and navigation of these contents.

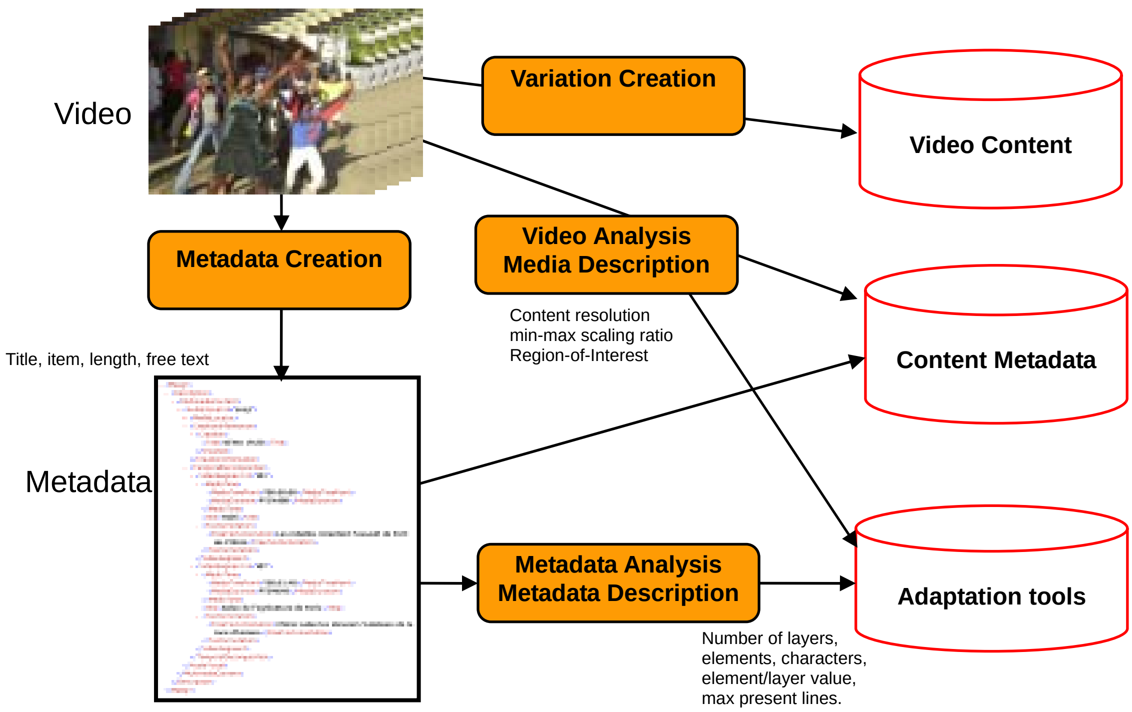


Figure 39. Preparation of the necessary descriptions

10.3. Preparation of the necessary contents and descriptions

As explained in 10.2, three types of data (video contents, content metadata and adaptation tools) needs to be prepared beforehand in this system.

1. Video content

The video content to be delivered is prepared beforehand. The video content can be adapted by transformation and/or variation selection. In adaptation by transformation, the content is adapted on-the-fly while adaptation by variation selection selects the most appropriate variation from those created beforehand. Those two approaches can be combined. For example, first the system selects the best variation and then this variation is transformed to fit more precisely. Variations of a single content with different resolution is necessary to be created beforehand to adapt the content by variation selection.

2. Content metadata

The content metadata describes the content. There are metadata used for adaptation process and metadata used to be presented to the consumer. Media description describes the media itself, like the resolution, format and duration of the video content. This metadata can be created automatically. Structural and semantic description describes the structure of the video and text annotation for each structured video segments. For example, title, free text annotation, and the duration for each segment. The title and free text annotation for the whole content are also described. If there are any variations of the video content, variation description describes the associations or relationships between different variations of multimedia content. Media descriptions are used for capability negotiation to determine how to adapt the video content. On the other hand, textual information like title and free text annotation of each segment and the whole content are used to be presented to the consumer.

3. Adaptation tools

The adaptation tools in this system provide hint information for guiding the adaptation process. There are tools for helping video content adaptation and also tools for helping metadata adaptation. For video content adaptation, we focus on spatial aspects because this effects directly to the screen size. For metadata adaptation, we describe the number of layers (e.g. program-topic-scene-shot…) , elements and characters included in the metadata user for presentation to enable dynamic control of the information to be presented to the consumer.

Video content adaptation tools:

- Minimum spatial scaling rate: This parameter is defined as a minimum rate that the content can be scaled. If the video content is resized too small, the involved objects cannot be perceived. Some minimum spatial scaling rate has to be defined so that the involved objects can be perceived.
- Maximum spatial scaling rate: This parameter is defined as a maximum rate that the content can be scaled. For example, when a content whose resolution is 320x240 has to be presented in a screen of 1024x768, it is not a good idea to enlarge more than three times the video content resolution because the perceptual image quality becomes very low. Some maximum spatial scaling rate needs to be defined (depends on the context).
- Value of each variation: If any variations are prepared, we defined a content value for each variation so that a video content transformed from the original source and a video content transformed from one of the variations can be compared using the same measure.

Metadata adaptation tools:

- Number of layers of the metadata for presentation.
- Number of elements of the metadata for presentation.
- Number of characters for each element.
- Value of each element.

```
...
<ccpp:component>
  <rdf:Description rdf:about="TerminalHardware">
    <rdf:type rdf:resource="HardwarePlatform" />
    <ccpp:pix-x>240</ccpp:pix-x>
    <ccpp:pix-y>320</ccpp:pix-y>
  ...
  </rdf:Description>
...
</ccpp:component>
```

```
...
<DIADescriptionUnit xmlns="urn:mpeg:mpeg21:2003:01-DIA-NS"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:schemaLocation="urn:mpeg:mpeg21:2003:01-DIA-NS UED.xsd"
 xsi:type="DisplayCapabilityType">
 <Mode>
   <Resolution horizontal="240" vertical="320" />
 ...
 </Mode>
...
</DIADescriptionUnit>
```

Figure 40. Example of a screen size description in pixels using CC/PP and MPEG-21 DIA

10.4. Description of the consumer side

For the consumer side, device capabilities and user preferences are described so that they can be used to guide the adaptation of the content presented to that device. We describe the screen size of the device and preferences related to browsing. These features can be described using both CC/PP and MPEG-21 DIA.

1. Screen size

For screen size adaptation, it is essential in the video content adaptation process to have the screen size in pixels to determine the most appropriate size of the video that can be displayed. It is also essential in metadata adaptation process to have the number of characters and lines that can be displayed in the screen to determine the most appropriate form of the metadata to be delivered. CC/PP attribute vocabulary for print and display is given as an example to describe these features.

A. The number of horizontal and vertical pixels that can be displayed.

The height and the width of the screen where the contents are delivered are described. In case the content is presented in a browser inside the screen, the size of the browser is described. **pix-x** and **pix-y** from CC/PP attribute vocabularies for display are used to describe this feature.

pix-x: (Value data type: Integer) The number of horizontal pixels that can be displayed.

pix-y: (Value data type: Integer) The number of vertical pixels that can be displayed.

B. The number of characters and number of lines of text that can be displayed.

The width of the character display and the number of lines of text that can be displayed are described. **charWidth** and **charHeight** from CC/PP are used to describe this feature.

charWidth: (Value data type: Integer) The width of the character display. For non-proportional font displays, the number of display cells. For proportional font displays, the width of the display in *ens* (where an *en* is the typographical unit that is the width of an *en*-dash/letter 'n').

charHeight: (Value data type: Integer) The number of lines of text that can be displayed (i.e. the display height in characters).

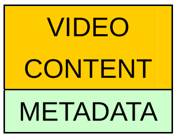


Figure 41. Example of layout preference (video : metadata = 2 : 1)

2. Browsing preferences

The browsing preferences are designed for presenting the content in a way that the consumer prefers. We consider two types of preferences; the layout structure of the presented contents, and the importance of the video content and metadata for the consumer. These data is not mandatory for the consumer, and if there is no preference, the system tries to find the best layout to provide the best browsing experience to the consumer.

A. Layout

Layout preference describes the preferred balance of video pane size and metadata pane size to be presented to the device, and the preferred position of the video content (left, right, top, bottom of the screen) to be displayed. For example, if the preferred balance of the video pane size and metadata pane size is 2:1, then two third of the screen is filled with the video content and the rest with metadata.

B. Importance

Importance describes how important the video content and the metadata are for the consumer. For example, if the user just wants to watch the video and metadata is just additional for him, video importance vs metadata importance could be assigned to 9:1 or 10:0.

10.5. Screen size adaptation engine

The screen size adaptation engine uses the device capabilities and user preferences of the consumer described in 10.4 as restrictions and the content descriptions and the adaptation tools described in 10.3 for optimizing the adaptation process. The adaptation engine dynamically and jointly determines the following three things; how to present the video being requested by the consumer, how to present its content metadata and how to balance the presentation of the video and its metadata; all in a way that best meets the screen size and the browsing preferences. Section 11 explains how to model this optimization problem.

Next, the video content and its metadata are transformed by applying the determined optimized way of adaptation. About the video content, the variation that is downscaled with the closest scaling ratio is selected and retransformed to the determined size. The metadata are scaled considering the tradeoff between the amount of the involved information and the ability to present them in the given screen size. The details for video content adaptation and metadata adaptation are given in section 12.

11. Joint adaptation of video content and its metadata

This section presents how to adapt jointly a video content and its metadata in a way that maximizes the content value for the user within a fixed screen size. There are three factors to be considered in content presentation to a specified screen size; how to present the video content, how to present its metadata, and how to balance the presentation of the video and its metadata. The most straightforward presentation way is to pre-define a fixed size of a video pane size and a metadata pane size, then adapt the video content in a way that the video content value to be presented in the video pane is optimized, and also adapt the content metadata of the requested video in a way that the value of the content metadata be maximized. However, as every video contents and every content metadata has their own data size, and as every user has different browsing preferences and devices with different browser size, a flexible presentation of the contents with its metadata is essential to provide the best experience to the user. In this section, we present an approach to find an optimal balance of the presentation area (pane) for the requested video content and for the metadata in a way that maximizes the total value of the content provided to the user.

11.1. Joint adaptation process for overall optimization

Figure 42 describes the joint adaptation process of a requested video content and its metadata. The adaptation process decision engine determines the pane layout that optimizes the balance between video content pane and metadata pane. It also decides the optimal way to adapt the video content and its metadata to the panes. This decision is jointly made by considering all the video descriptions, metadata descriptions, the screen size and the browsing preferences. After determining the optimal pane layout and the optimal content adaptation parameters, the video content and its metadata are transformed based on the determined parameters. The transformed video content and metadata are presented over the determined pane layout.

11.2. Layout control for balancing video and metadata presentation

There are two aspects for determining the layout of how to present the video content and its metadata; the **pane layout** (horizontal, vertical, etc…) and the **size of each pane** (video pane, metadata pane).

There are many possibilities in how to present the video and the metadata in a fixed screen size. For example, locate the video content and fill all the rest of the screen with text metadata, fill thumbnails in redundant spaces, overlay some text over the video content to maximize the video content size, determine the area for presenting video and metadata not only in rectangular form but also in any form like triangular, circular, and so on.

Here we consider the four possibilities shown in Figure 43 to simplify the problem; the video, the metadata to be located up, down, left, or right. The screen is simply divided into two parts, horizontally or vertically, and one part is assigned for video and the other for metadata.

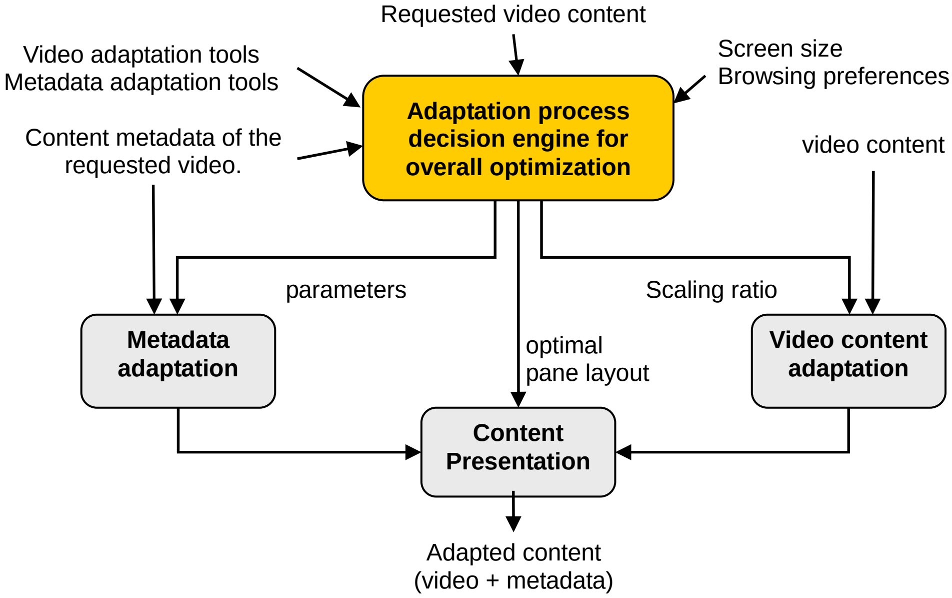


Figure 42. Joint adaptation process of a video content and its metadata

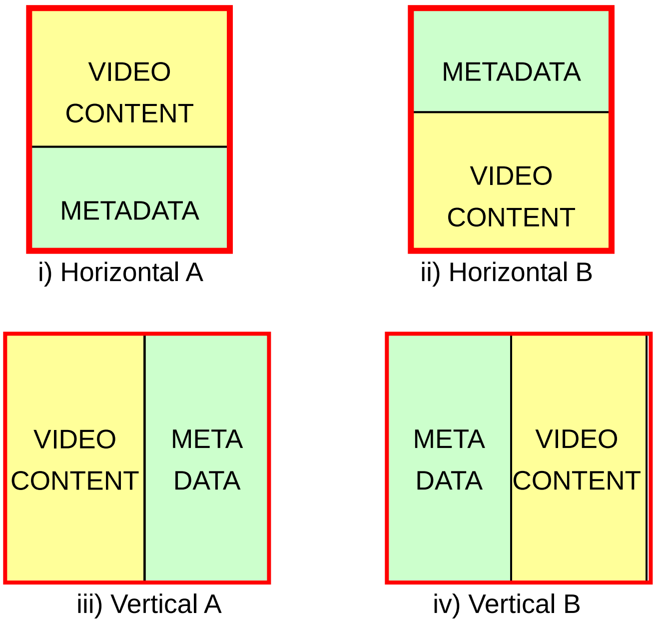


Figure 43. Pane layout patterns for presenting a video content and its metadata.

For layout determination, there are two things the adaptation engine decides. The first one is that if the screen should be divided horizontally or divided vertically, and the second one is the sizes of each divided area. We just focus on “Horizontal A” case and “Vertical A” case in Figure 43 since the area size of each modality remains the same with “Horizontal B” case and “Vertical B” case. The difference of them is just if the video are located up or down, left or right. We cover this difference by considering the user preference or let the user choose the preferred layout after determining the better case between the former two cases “Horizontal A” and “Vertical A”.

11.3. Total content value function

To optimize the content adaptation process, a quality metric to evaluate how well the content is adapted meeting the constraints on both the provider side (e.g. restrictions on scaling rate) and the consumer side (e.g. device capabilities, user preferences) is emerging. We define a “value” of the total adapted content to evaluate the fidelity of the overall adapted content compared with the original content. Thus, the optimization process can be described as determining the way of adaptation that maximizes the total content value meeting the given constraints.

Several approaches to define the value of the adapted content have been proposed. Mohan [Mohan99] proposed a general value-resource framework to optimize the selection of a best variation from multiple modalities and multiple resolutions derived from a single content. He introduced a subjective measure of fidelity called “value” defined as the perceived value of the selected variation divided by the value of the original content. Using this “value”, he extended the rate-distortion theory, which deals with the allocation of bits and the distortion of the source, to a general value-resource framework that deals with the different client resources and the “value” of the selected content. The main challenge remains in how to define a specific function to describe the relation between this value and the given resource. Lee [Lee01] applied this value-resource framework in image transcoding. He proposed an image resizing method using this framework that tries to fit the image with the client display size considering the object importance inside the image. Importance values are added beforehand to each block inside the image, and the resizing process is performed by optimizing the combination of cropping and scaling so that the value of the resized image is maximized. Xie [Xie04] introduces a conceptual model to optimize a presentation of a content with multiple variations (information objects) by defining an importance value, a minimal perceptible size to denote the minimal allowable spatial area and an alternative as a substitute of the original content for each variation. A function to determine the way to balance the adaptation of multiple variations still remains as a challenge.

We propose a total content value function to balance the adaptation of the video content and its metadata under the given screen size and browsing preferences. As described in 11.2, it is necessary to determine dynamically how to divide the screen for balancing the presentation of the video content and its metadata. There are two aspects in dividing the screen, the size and the layout. The content value function based on the pane size is described in section 11.3.1. Section 11.3.2 gives how to combine the layout aspect in this function so that the total content value considering both aspects can be evaluated with a single function.

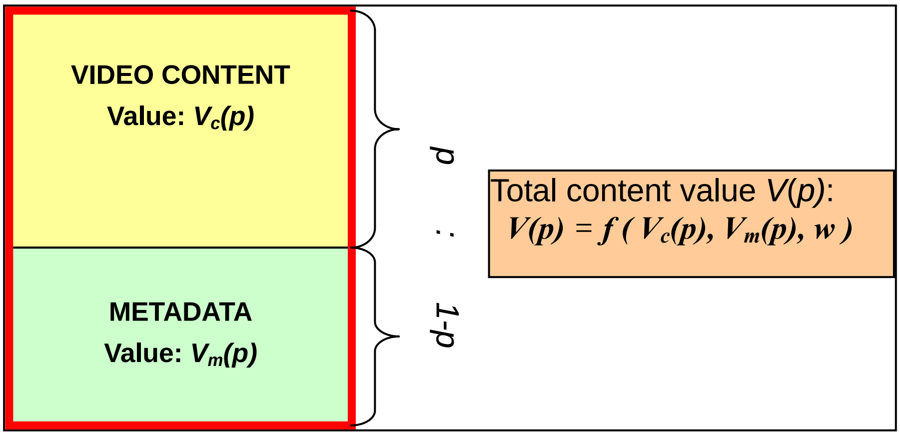


Figure 44. Overall content value definition

11.3.1. Value function considering the size of each pane

For expressing the balance of the size of each pane, we define a variable p as a normalized value of the video content pane size (Figure 44). This value is normalized to be between 0 (whole screen is for metadata) and 1 (whole screen is for video). Thus the balance of the video content pane and metadata pane can be expressed as $p:1-p$ in both horizontal and vertical layout case.

Once the pane size to present the video and that of the metadata are determined, the total content value can be calculated by optimizing the content value of each modality in each given pane size. This means that the value of the total content V can be defined as a function of p.

There are three parameters that affect the total content value. The video content value, the metadata value, and the preferred balance of them are the three parameters. The total content value function $V(p)$ is modeled as follows;

$$V(p) = f(V_c(p), V_m(p), w) \qquad (0 < p \le 1) \tag{1}$$

where

$V_c(p)$: value of the adapted video content
$V_m(p)$:value of the adapted metadata.
w: importance of the video content (normalized) .

The total content function $V(p)$ evaluates the balance between the adapted video content and the adapted metadata and also takes into account the subjective importance of each modality for the consumer. Considering the video content value and metadata value, as both the video content pane size and the metadata pane size can be calculated from the screen size and the given p, the value of both modalities can also be expressed as a function of p. It should be noted that any definition of the adapted video content and metadata value function is applicable.

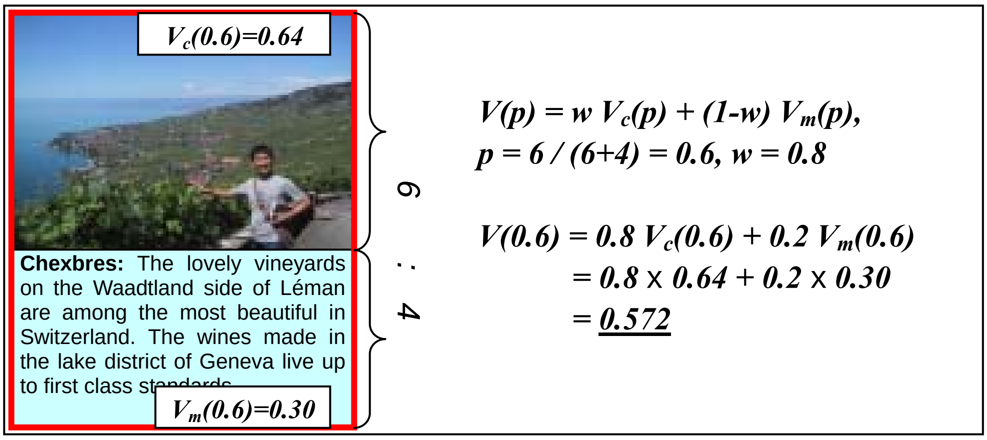


Figure 45. An example for calculating total content value

Our approach to optimize the adaptation for the video content to calculate $V_c(p)$ is described in detail in section 12.2. The optimization process of metadata adaptation to calculate $V_m(p)$ is described in section 12.4. w is a normalized value of the video importance which is given by the user as browsing preferences as explained in 10.4. The balance between the importance on video content and on metadata can be expressed as $w : 1\text{-}w\ (0<w<1)$. When the importance value is not given by the user, $w=0.5$ is assigned as a default value.

The definition of function f, how to fuse two different value dimensions into one value remains as a big challenge. As the balance of the video pane and the metadata pane would cause effects to the human visual system, subjective experiments are needed to define a precise function f. A typical experiment would be preparing various screen size, and prepare various combinations of video-metadata balance within each screen size, and let people score every combination with a subjective value.

In this paper, we apply the following equation which seems logical and simple;

$$\begin{aligned} &V(p) = f(V_c(p), V_m(p), w) = w \cdot V_c(p) + (1-w) \cdot V_m(p) \\ &0 < p \le 1 \\ &0 < w \le 1 \end{aligned} \quad . \qquad \textbf{(2)}$$

Figure 45 shows an example using the equation (2) for calculating the total content value where p = 0.6 (video pane: metadata pane = 6 : 4) and w = 0.8. The maximum video content value for p=0.6 is 0.64 and the maximum metadata value for p=0.6 is 0.30. The total content value is calculated as a weighted sum of both modalities, which results in 0.57.

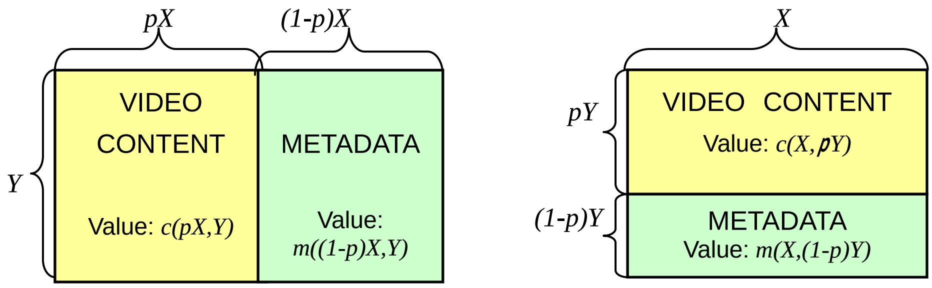


Figure 46. Video content pane size and metadata pane size for each layout

11.3.2. Adding the pane layout to the value function

The other challenge is how to determine the pane layout. As described in 11.2, there are two possibilities in layout determination, dividing the screen horizontally or vertically.

Our approach is to evaluate both possibilities and select the one with higher total content value. The total content value can be determined by calculating the maximum total content value for both horizontal and vertical layout and selecting the one with higher content value. The total content value $V(p)$ becomes:

$$V(p) = \max\{V_H(p), V_V(p)\} \quad \textbf{(3)}$$

where

$V_H(p)$: Total content value in horizontal layout.

$V_V(p)$: Total content value in vertical layout.

The function based on the pane size and the function based on the pane layout are combined in a single one to evaluate the total content value independent of the layout. Figure 46 illustrates the video content pane size and metadata pane size for both horizontal and vertical layout. By applying the formula (1), the total content value for both layout cases, $V_H(p)$ and $V_V(p)$, can be modeled as;

$$\begin{aligned} V_H(p) &= f(c(pX, Y), m((1-p)X, Y), w) \\ V_V(p) &= f(c(X, pY), m(X, (1-p)Y), w) \end{aligned} \quad \textbf{(4)}$$

where *(X, Y)*, *c(x, y)*, *m(x, y)* are defined as follows:

(X, Y): Screen size.
c(x, y): Maximum video content value in pane size *(x, y)*. Details in 12.2
m(x, y): Maximum metadata value in pane size *(x, y)*. Details in 12.4.

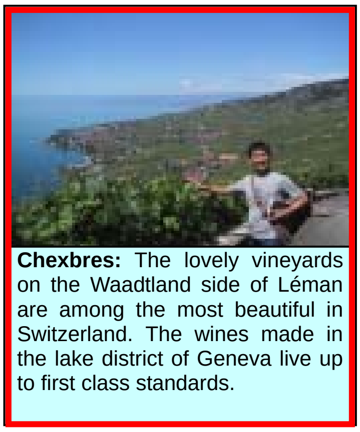
	Horizontal layout (p=0.6)	Vertical layout (p=0.6)
		
Video content value	$V_c(0.6) = 0.64$	$V_c(0.6) = 0.15$
Metadata value	$V_m(0.6) = 0.30$	$V_m\ (0.6) = 0.30$
Total content value	$\mathbf{V_H(0.6) = 0.572}$	$V_V\ (0.6) = 0.18$

Figure 47. Example of layout decision by comparing total content value in both horizontal layout case and vertical layout case

The value of $V_c(p)$ and $V_m(p)$ can be determined using the following conditions;

$$
\begin{aligned}
&\mathit{if}\ (V(p) == V_H(p)) \\
&\{V_c(p), V_m(p)\} = \{c(pX, Y), m((1-p)X, Y)\} \\
&\mathit{else} \\
&\{V_c(p), V_m(p)\} = \{c(X, pY), m(X, (1-p)Y)\}
\end{aligned}
\qquad \textbf{(5)}
$$

We define the total content value for both layout cases by applying the formula (2) to the function (4). The weighted sum of the value of both panes are used to form the total content value for *p*. The total content value is defined as follows;

$$
\begin{aligned}
V(p) &= \max\{V_H(p), V_V(p)\} \\
V_H(p) &= w \cdot c(pX, Y) + (1-w) \cdot m((1-p)X, Y) \\
V_V(p) &= w \cdot c(X, pY) + (1-w) \cdot m(X, (1-p)Y)
\end{aligned}
\qquad \textbf{(6)}
$$

Figure 47 shows an example for layout determination by comparing the total content value in both horizontal layout case and vertical layout case. In horizontal layout case, the video content and its metadata are located horizontally. In this example, the video content itself fits almost perfectly the area for the video content and results in a video content value of 0.64. On the other hand, in vertical layout, there are a lot of blank space due to the difference of the aspect ratio of the video content and that of its presentable area. Therefore, the video content value is very low, 0.15, in this case. The difference in video content value for both layout results in having a large difference in total content value. In this case, the horizontal layout is used as it has larger value.

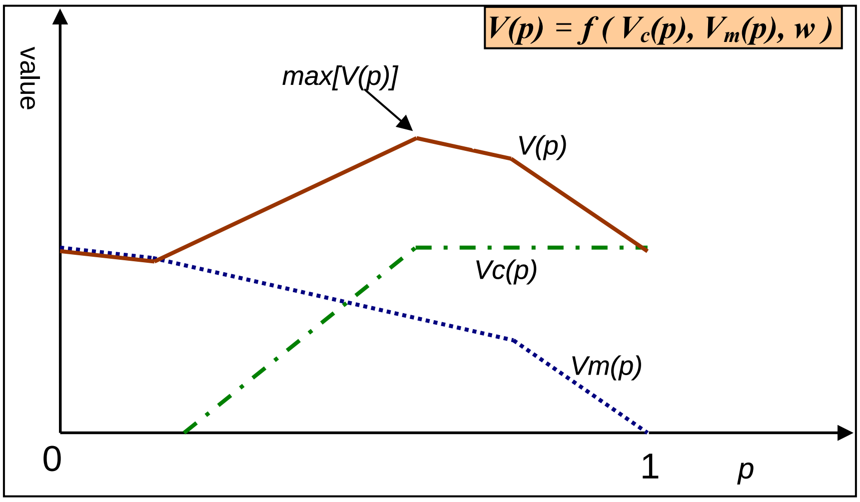


Figure 48. Determination process of the *p* value that maximizes the total content value

11.3.3. Optimization of the total content value function

Considering all the discussions above, we can regard the optimization problem of content adaptation as a problem to **find a value *p* that maximizes the total content value**. Once this *p* is determined, the pane layout, the size of each pane, the way how to optimize the video content value and the metadata value are determined at the same time. This problem is given by the following:

$$\hat{p} = \arg\max_p [V(p)] \qquad (0 < p \leq 1) \tag{7}$$

Figure 48 illustrates an example of the determination process of *p* that maximizes the total content value. For each *p*, the video content value $V_c(p)$ and the metadata value $V_m(p)$ is maximized, respectively. As a result, a ***V-p curve*** (content value - layout balance) can be drawn for both $V_c(p)$ and $V_m(p)$. Then, the total content value for any *p* is calculated by using the equation (1) which also appears in the *V-p curve*. The optimal layout is determined by selecting the *p* value which maximizes the total content value *V(p)*.

By applying the equation (6) as $V(p)$, the optimization problem considering both the screen size and the browsing preferences can be described as shown in Figure 49.

$$\hat{p} = \arg\max_p [\max\{w \cdot c(pX, Y) + (1-w) \cdot m((1-p)X, Y),$$
$$w \cdot c(X, pY) + (1-w) \cdot m(X, (1-p)Y)\}] \tag{8}$$

(X, Y): Screen size,
$c(x, y)$: Maximum video content value in pane size (x, y),
$m(x, y)$: Maximum metadata value in pane size (x, y) .

Figure 49. Equation to optimize the total content value

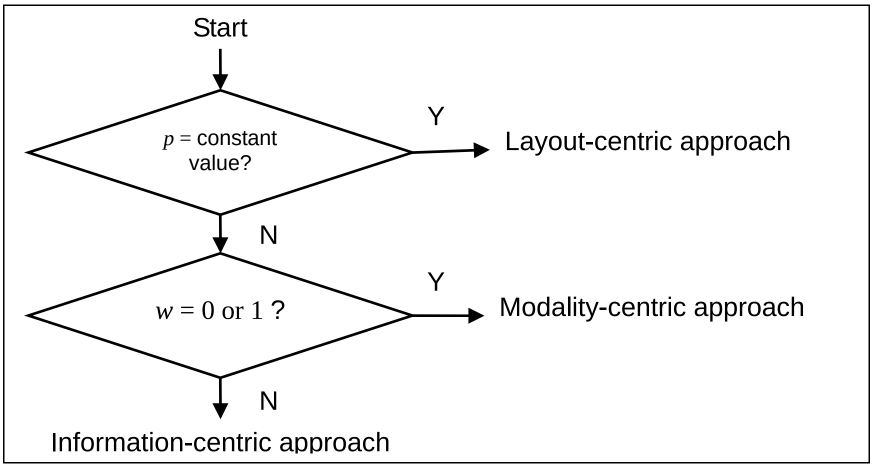


Figure 50. Decision of adaptation process approach

11.4. Optimizing content adaptation considering browsing preferences

Considering the browsing preferences, we categorize the approaches for optimizing content adaptation into three types. These approaches are based on which of the following are most important for the consumer; the total amount of the presented information, the preferred modality (video or text), or the layout of the whole content. We call them **a) information-centric approach**, **b) modality-centric approach**, and **c) layout-centric approach**, respectively. The approach applied for the adaptation process is automatically determined by looking the browsing preferences, to be specific, the value p and w. Figure 50 shows its determination procedure.

a) Information-centric approach. (p = variable, $0 < w < 1$)

Information-centric approach is applied in case the user wants to have as much valuable information as possible regardless of the modality. In this approach, the video content and the metadata are balanced adaptively in a way that maximizes information throughput. This approach is applied when value p is flexible and value w ranges $0<w<1$. Figure 51 shows some examples obtained by this approach. Details are presented in 11.4.1.

b) Modality-centric approach. (p = variable, w = 0 or 1)

Modality-centric approach is applied when the user wants to browse mainly a single modality and the other modality is just optional for him/her. The typical case for this approach would be the case what is important for the user is just to watch a video content and he/she doesn’t care about metadata at all. This approach is applied when w = 0 or w = 1, and p is flexible. Figure 52 shows some examples obtained by this approach. Details are given in 11.4.2.

c) Layout-centric approach. (p = constant)

Layout-centric approach is applied when the user wants to have video and metadata presented with a fixed balance and/or in a fixed position. A typical example would be when he/she feels relaxed when metadata is located under the video, or he/she always like to have the screen with two third with the video content and the rest with metadata. This approach is applied when a constant value p is given. Figure 53 shows examples with this approach. Details are in 11.4.3.

Figure 51. An example of adaptation by "information-centric approach": this approach tries to provide as much information in total as possible.

Figure 52. An example of adaptation by "modality-centric approach": The video size is always maximized.

Figure 53. An example of adaptation by "layout-centric approach": This user prefers the video content and its metadata area size to have always balanced 1 : 1 ratio.

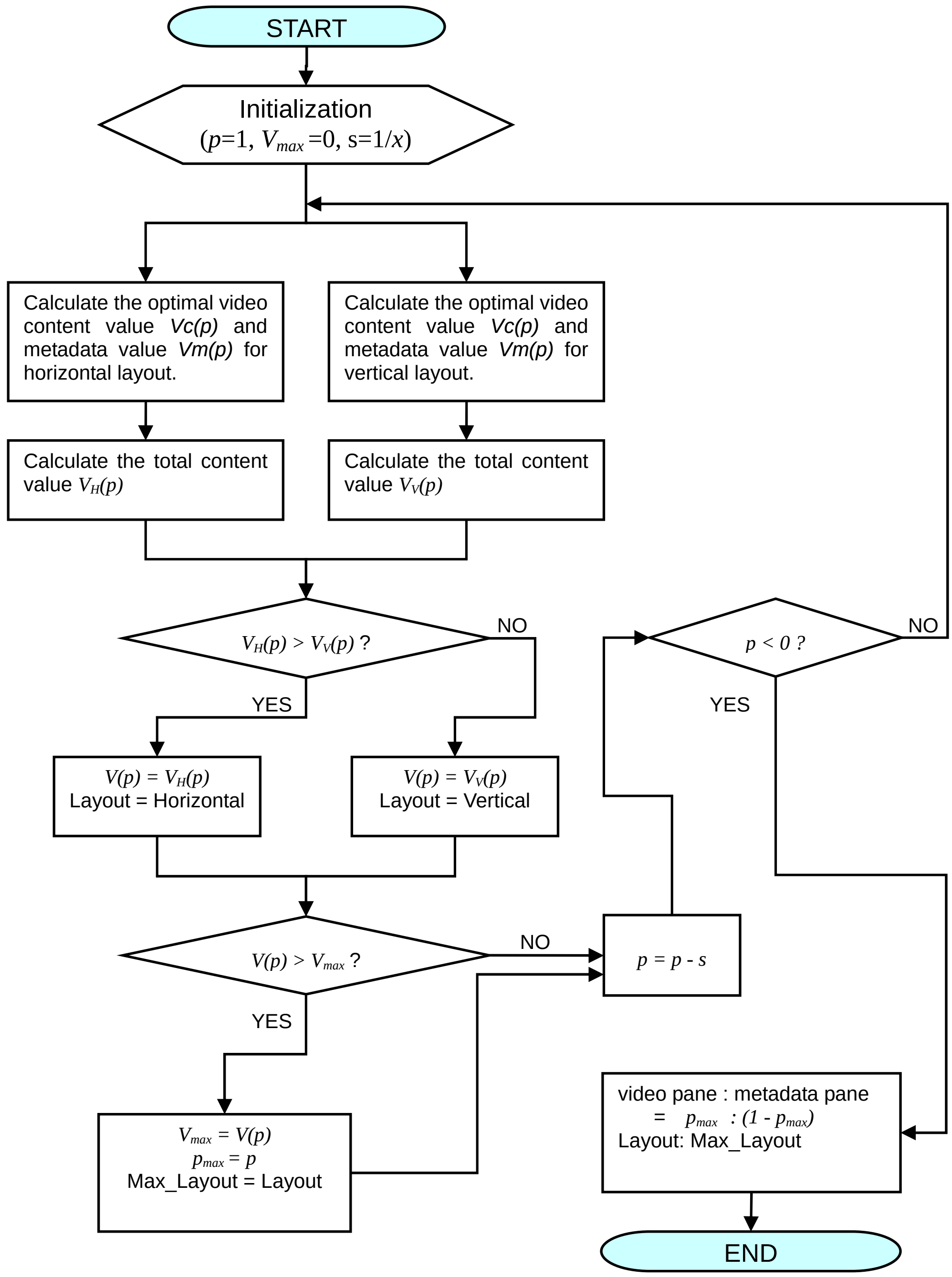


Figure 54. Optimization process of information-centric approach

11.4.1. Information-centric approach

The total content value is optimized in information-centric approach to provide as much information as possible to the consumer. In this approach, we try to find the value p that fulfills function (8) as described in 11.3. Figure 54 describes the detailed algorithm to find the optimal layout and value p. The total content value for every p is calculated and the value p which maximizes the total content value is selected as the optimal balance between video and metadata pane. The step size *s* for reducing *p* can be determined considering the available computational power for adaptation.

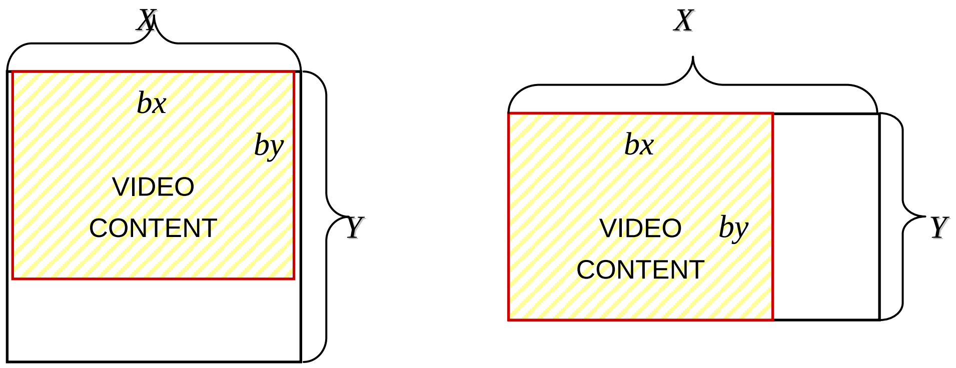


Figure 55. Video-centric adaptation approach

11.4.2. Modality-centric approach

In modality-centric approach, the layout is determined in a way that maximizes value the modality the user is interested. When $w=0$, which means that the user is only interested in metadata, the video is just ignored and the entire screen is used to present metadata. When $w=1$, which means that the user is only interested in video, the video content size (width x height) is maximized to fit the whole screen. In case there is enough room left in the screen after presenting the adapted video content, metadata is adapted to this area of the screen. If the area being left is smaller than a certain amount, the metadata is just ignored due to display capability. It should be noted that the aspect ratio of the adapted video content is always kept the same.

We try to maximize the video content size while keeping the video content size within the screen. If $w=1$ is applied in equation (8), this problem can be expressed as follows;

$$\hat{p} = \arg\max_{p}[\max\{c(pX, Y), c(X, pY)\}] \qquad \textbf{(9)}$$

However, as the number of possible p that maximizes the video content size is very small, we directly compare all the possibilities for faster operation instead of using function (9). Figure 55 shows the two possibilities that could maximize the video content size inside the screen. The first case is that the width of the adapted video content is identical to the screen width X. The second case is that the height of the adapted video content is identical to the screen height Y. We determine the layout and video content scaling ratio b in a following way;

__b = min (b1, b2)__   *where* $b1 = X / x$ *and* $b2 = Y / y$
if (b > bmin)
   __if (b == b1)   vertical layout.__
   __if (b == b2)   horizontal layout.__
if (b > bmax) b = bmax;
if (b < bmin) b = 0

Parameters:

b:	Video content scaling ratio.
(X, Y):	Width and height of the screen.
(x, y):	Width and height of the original video content.
bmax, bmin:	Maximum, minimum allowed scaling ratio of video.

Figure 56. Layout determination in video-centric adaptation approach

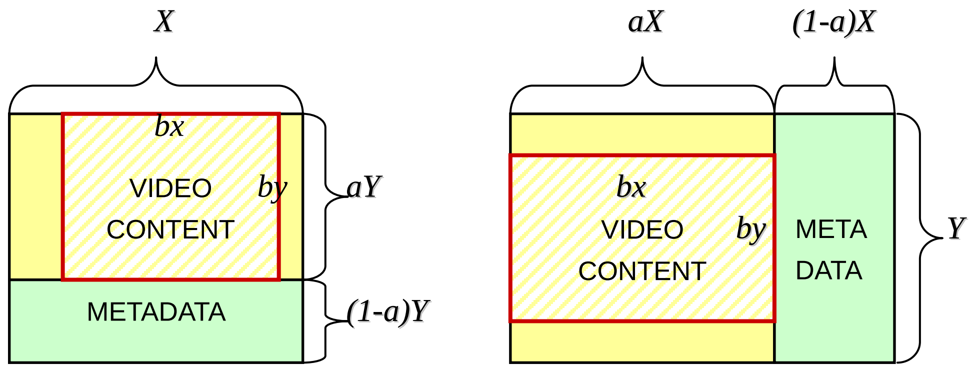


Figure 57. Content adaptation for fixed layout balance (always keep video : metadata = a : 1-a, a = const.)

11.4.3. Layout-centric approach

In layout-centric approach, the balance of the size of both panes is already determined by the user *(p=a, a=const.)*. Therefore, we try to maximize separately the video content value and metadata value within the given pane size for each modality. If the user prefers a fixed pane layout pattern, for example, he/she always wants "Horizontal A" pattern regardless of the screen size and its aspect ratio, then the adaptation process simply applies the preferred layout pattern and the video content and metadata are optimized in each pane, respectively.

The optimization procedure is different in case the user wants a fixed balance of content presentation but prefers to have the content presented in a flexible way instead of a fixed pane layout. For example, when the user has a screen whose height is much larger than its width, the user has the contents presented in horizontal layout. On the other hand, the contents could be presented in vertical layout when the width is much larger than the height. Thus, we can make the most of the user's display capabilities. In this case, there are two possibilities for layout determination, therefore, we calculate the maximum total content value for both horizontal and vertical layout for the given pane size and selecting the one with higher content value as given in equation (3).

Figure 57 shows the two possible layout patterns for the pre-defined layout balance. There are two methods for determining the layout pattern and the scaling size of the video content for fixed pane balance. The method differs depending on w value, $0<w<1$ case and $w=0,1$ case. Same principles as information-centric approach are applied for the former and the other as modality-centric approach. The methods are described in the following.

<u>**Parameters:**</u>

a:	Balance of video content and metadata presentation area given by the user (video : metadata = a : 1-a. a = constant value).
b:	Video content scaling ratio.
(X, Y):	Width and height of the screen.
(x, y):	Width and height of the original video content.
$bmax$, $bmin$:	Maximum, minimum allowed scaling ratio of video content.

$0<w<1$ case:

1. Maximize both video content and metadata independently in each pane.
2. Calculate the total content value for each pane layout pattern.
   Horizontal layout: $V_H(a) = f(c(aX, Y), m((1-a)X, Y), w) = wc(aX, Y) + (1-w)\, m((1-a)X, Y)$
   Vertical layout: $V_V(a) = f(c(X,aY), m(X, (1-a)Y), w) = wc(X, aY) + (1-w)\, m(X, (1-a)Y)$
3. Adopt the one with higher content value.
   $V_H(a) > V_V(a)$: Horizontal layout.
   $V_V(a) > V_H(a)$: Vertical layout.

w=1 case:

For $w=1$ case, we decide horizontal or vertical layout considering the maximum video content size that can be displayed in each layout. For fast calculation, we determine the layout and the video content scaling ratio b in a following way;

```
b5 = min (b1, b2)    where   b1 = X / x    and   b2 = aY / y
b6 = min (b3, b4)    where   b3 = aX / x   and   b4 = Y / y
b = max (b5, b6)

if (b > bmax) b = bmax;
if (b > bmin)
   if ( b == b5)   vertical layout.
   if ( b == b6)   horizontal layout.
if (b < bmin) b = 0
```

Figure 58. Layout determination in layout-centric approach.

12. Content adaptation for each modality

This section presents how a video content and its metadata are adapted in a given area. For video content adaptation, the video content resolution is maximized to fit the video pane without changing its aspect ratio. Metadata is adapted considering the tradeoff between the amount of the involved information and the ability to present them in the given metadata pane. Section 12.1 - 12.2 describes how to define the video content value and its adaptation process. The definition of metadata value and its adaptation process are presented in section 12.3 - 12.4.

12.1. Video content adaptation

Our approach for video content adaptation is to deliver the video context as is, in other word, as the content creator or service provider desires. Therefore, we simply adapt the spatial resolution of the video content keeping its aspect ratio to the video pane instead of changing the video context, for example, changing the modality (video to key-frames, video to text, etc) of the content, zooming primary objects, or selecting adaptively the objects to be included inside the video. Note, temporal resolution is also important but in this paper we don't consider it as it does not effect in screen size adaptation.

Figure 59 shows the video content adaptation process and its parameters. The original video content is scaled to fit the given video pane. The adaptation engine determines the scaling ratio that maximizes the video content size. It should be noted that the resolution of the adapted video content must not exceed the size of the video pane.

12.2. Modeling video content value

As described in section 11, it is necessary to define the value of the adapted video content to determine the best balance between video content adaptation and its metadata adaptation. From the equation (5) in section 11.3.2, it is obvious that there are two possible video content values $V_c(p)$ corresponding to the given *p*. The two possibilities are as follows;

$$\begin{aligned} V_c(p) &= c(pX,\ Y) \quad (V_H(p) > V_V(p)) \\ V_c(p) &= c(X,\ pY) \quad (V_H(p) < V_V(p)) \end{aligned} \qquad \textbf{(10)}$$

To determine the video content value $V_c(p)$, function $c(x, y)$ that indicates the video content value corresponding to the given pane size (x,y), has to be defined. Section 12.2.1 describes how to define the function $c(x, y)$ and some possible extensions are introduced in 12.2.2.

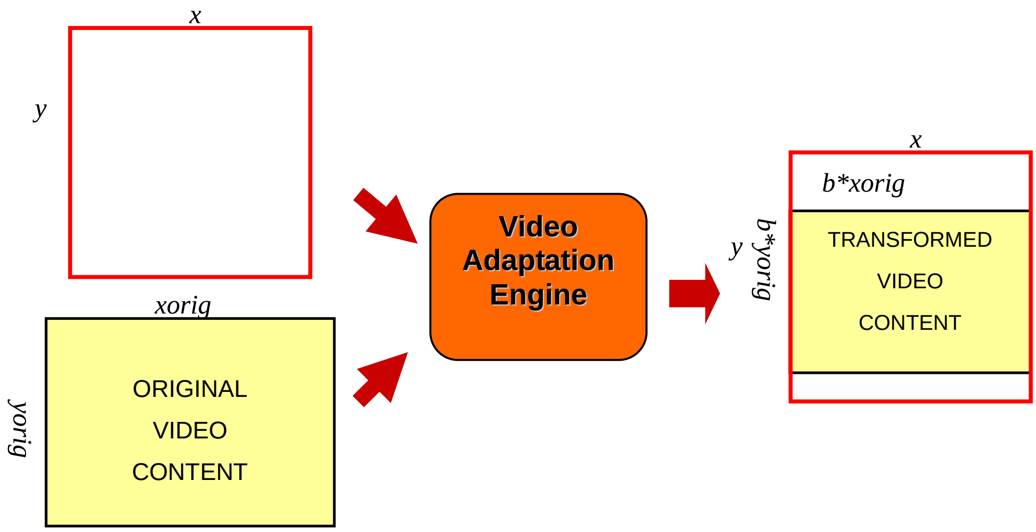


Figure 59. Process and parameters of video content adaptation

12.2.1. Video content value

The adaptation engine determines the scaling ratio that maximizes the video content size. The video content value is affected by one single factor, the scaling ratio. The bigger the image size is, the larger the content value has. Thus, the video content value c on a specific video pane size can be defined as a function f_c of scaling ratio m as follows;

$$
\begin{aligned}
c &= f_c\,(m) && (b_{min} < m < b_{max}) \\
c &= 0 && (m < b_{min}) \\
c &= f_c\,(b_{max}) && (m > b_{max})
\end{aligned}
\qquad \textbf{(11)}
$$

where b_{max}, b_{min} are the maximum and minimum allowed scaling ratio of the video content, respectively. Thus, we can describe the video content value in video pane of size (x, y) as presented in Figure 60.

Video content value c(x, y) in video pane of size (x, y)

$c(x, y) = f_c\,(b)$

$b = \min\,(b1, b2)$ where $b1 = x\,/\,x_{orig}$ and $b2 = y\,/\,y_{orig}$

$\text{if } (b > b_{max})\ b = b_{max}\ ;$

$\text{if } (b < b_{max})\ b = 0$

Parameters:

b:	Video content scaling ratio.
(x, y):	Width and height of the video pane.
(x_{orig}, y_{orig}):	Width and height of the original video content.
b_{max}, b_{min}:	Maximum, minimum allowed scaling ratio of video.

Figure 60. Calculation of video content value in the given video pane

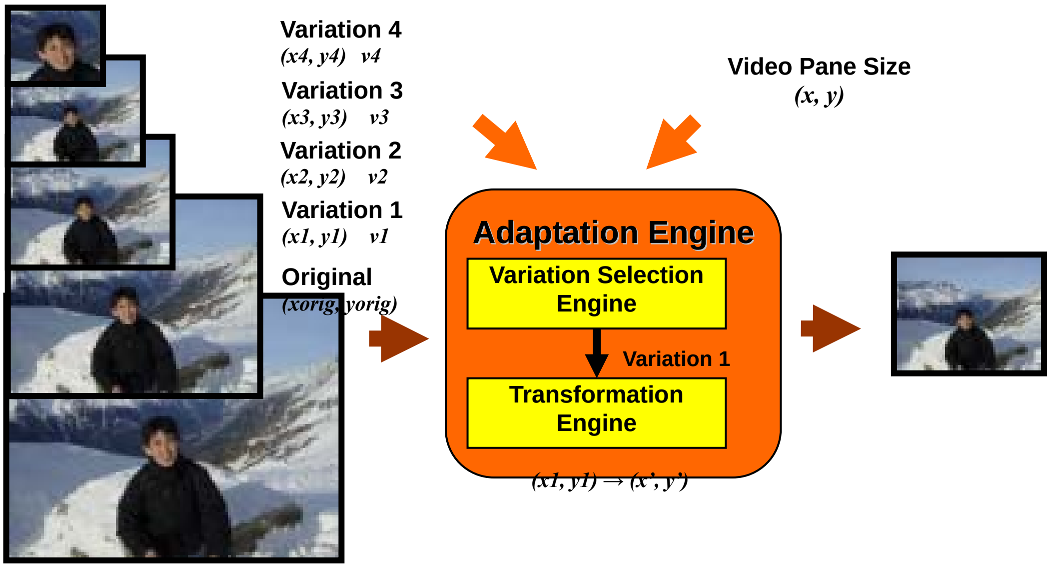


Figure 61. Video content adaptation process with variations

To define the relation between the scaling rate and the perceived importance, that is the definition of function f_c , remains as a big challenge. As the size of the content could cause effects to the human visual system depending on many factors like the context, motion, texture, color, etc., subjective experiments are needed to define a precise function f_c. A typical experiment would be preparing various contents scaled in various sizes, and let people score every variation with a subjective value.

In this paper, we use the scaling rate itself as a video content value which can be considered as a value that reflects the fidelity of the video content.

12.2.2. Video content value with variations

To adapt better the video content or to accelerate the video content scaling process, variations of the original content with different resolution can be prepared beforehand. Figure 61 illustrates the video content adaptation process with variations. The adaptation engine consists of a variation selection engine and transformation engine. The variation could be just a downscaled version keeping the same aspect ratio, and also could be a variation with different aspect ratio by spatially cropping every video frames.

In case variations of a single content with just different resolution are prepared, the adaptation engine selects the variation which has the closest resolution and the selected variation is transformed to fit precisely the video pane. The value for each transformed variation can be calculated in the same way as described in Figure 60.

On the other hand, if there are variations with both same aspect ratio and different aspect ratio as the original content, the approach is different. In addition to the transformed video content value, we also have to consider which variation could fit best the video pane in terms of aspect ratio. Therefore, the value for each variation with different aspect ratio has to be calculated in order to select the best one. Values, maximum and minimum scaling ratio have to be assigned for each variation with different aspect ratio. In this case, the video content value $c(x, y)$ in video pane of size (x, y) is calculated in the following way;

Video content value $c(x, y)$ in video pane of size (x, y):

$$c(x, y) = \max \{ c_1 (x, y), c_2 (x, y), \ldots\ldots c_i (x, y), \ldots\ldots, c_j (x, y)\}$$

$$c_i (x, y) = v_i \cdot f_c (b), \quad b = \min (b1, b2) \quad \text{where} \quad b1 = x / x_i \quad \text{and} \quad b2 = y / y_i$$

$$\text{if } (b > b_{imax}) \quad b = b_{imax}$$

$$\text{if } (b < b_{imax}) \quad b = 0$$

Parameters:

(x, y)	Width and height of the video pane.
j	Number of variations with different aspect ratio.
$c_i (x, y)$	Video content value for variation i in video pane of size (x, y)
v_i	Video content value of the variation (before transformation).
b	Scaling ratio of the variation.
(x_i, y_i)	Width and height of the variation i.
b_{imax}, b_{imin}	Maximum, minimum allowed scaling ratio of the variation i.

Figure 62. Calculation of video content value with variations

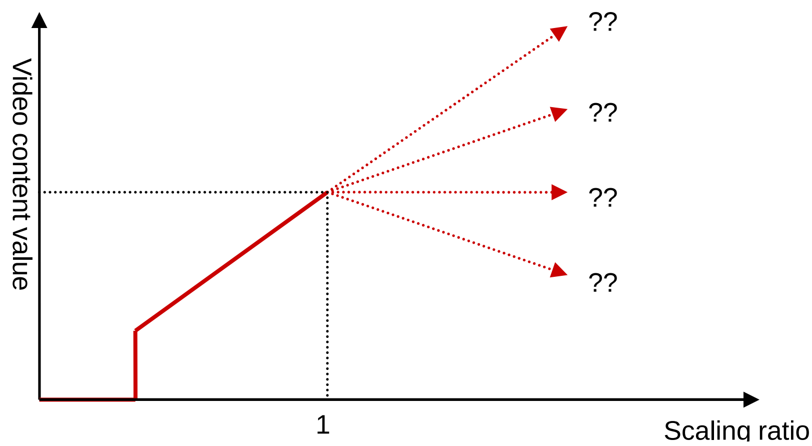


Figure 63. Relation between scaling ratio and video content value for upsampling

12.2.3. Additional parameters

Future work would be adding parameters like region-of-interest for frame/shot/scene of the video content, priorities on specific regions like face region, text region, primary objects, and relation between low-level features (texture, Motion activity, color, ….) , scaling rate and perceived content value. Value definition on enlarging the image is also necessary. If there is still space left in the video pane after putting the adapted video content, it is also helpful to fill this space with other information like thumbnails of key-frames, text information and so on. It is also important to define the value of the video contents that are upsampled (scaling ratio>1). The content value becomes higher in terms of visibility, becomes lower in terms of image quality, and remains the same in terms of fidelity. Therefore, it is necessary to balance these three factors.

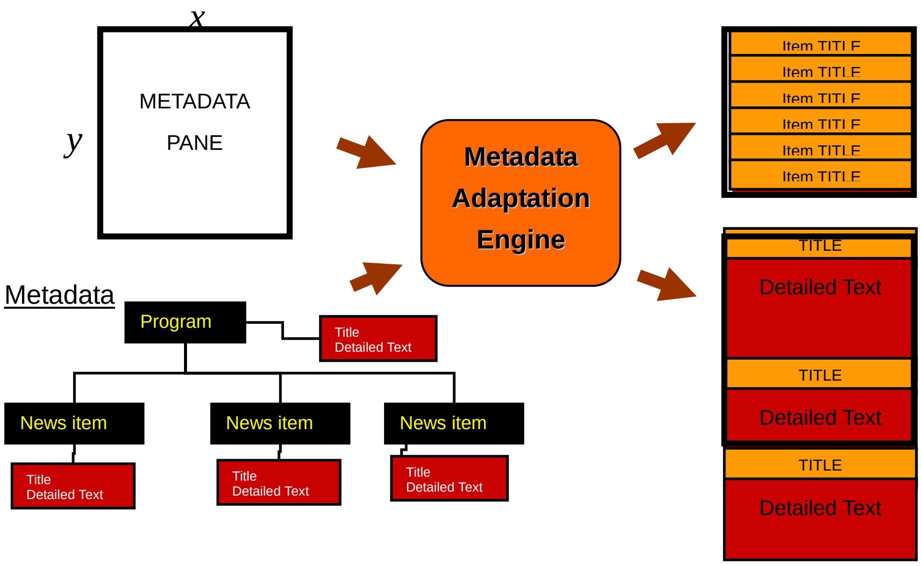


Figure 64. Example of metadata adaptation process

12.3. Metadata adaptation

Metadata adaptation is designed to decide how to scale and present the given metadata into a certain metadata pane size in an optimal way. Two factors must be balanced in metadata adaptation: the amount of information to be presented and how easily the presented information can be browsed in the given metadata pane. These two factors are contradictory. A large amount of information would be difficult to be browsed in limited metadata panes. Too much information in a small screen annoys the user. On the other hand, for easy browsing, only limited information can be presented in many cases. It is best if it is possible to display as much information as possible using as little lines as possible. Metadata adaptation includes the process to find the best tradeoff between these two factors.

Figure 64 shows an example of metadata adaptation process. A hierarchically structured metadata of a news program with several news items is prepared. Given this metadata and a metadata pane of a fixed size, the adaptation engine determines which information to present within the metadata in a way that optimally fits the metadata pane.

A practical way for reducing the required metadata pane size for presenting structured metadata, such as metadata of TV news programs and e-learning contents, is controlling levels of layers to be presented. Once a level of the layers to be kept is determined, the metadata is scaled to this level and all entities of the elements below that level is removed. Taking a TV news program as an example, the adaptation engine can determine if just metadata of the program, or those of each news item, or those of each shot are presented depending on the screen size.

Controlling the amount of text information for each element entity is another practical solution. For example, only the program title or the title of the current browsing news topic could be given in an extra-small screen. In a small screen, titles for all news topics can be given. The bigger the screen becomes, the more sentences for each topic can be presented. Whole metadata can be presented for a big screen.

One other possible extension is controlling the font size of the displayed metadata considering the pane size and the involved information. For example, a small font can be applied for giving more information to the user. On the other hand, large fonts are useful for the user to read easily the presented information.

Section 12.4 describes the definition of metadata value. This definition plays a key role to evaluate the quality of adaptation in determining the optimal way of adaptation.

12.4. Modeling metadata value

To determine the best balance between video content adaptation and its metadata adaptation, it is also necessary to define the value of the adapted metadata in addition to the video content value. Two possible metadata values $V_m(p)$ corresponding to the given *p* derived from the equation (5) in section 11.3.2 are as follows;

$$V_m(p) = m((1\text{-}p)X,\ Y) \quad (V_H(p) > V_V(p))$$
$$V_m(p) = m(X,\ (1\text{-}p)Y) \quad (V_H(p) < V_V(p)) \qquad \textbf{(12)}$$

To determine the metadata value $V_m(p)$, we have to determine the function $m(x, y)$ that indicates the adapted metadata value corresponding to the given pane size (x,y).

12.4.1. Metadata value

To calculate the adapted metadata value $m(x, y)$, we balance the amount of information to be presented and how easily the presented information can be browsed in the given metadata pane. To be concrete, we define the following two values that represent these two factors;

1. **Information value V_I (0-1)**: How much information the user obtains. Details given in 12.4.2.
2. **Presentation value V_P (0-1)**: How easily the user can view all the presented information. Details given in 12.4.3.

The more detailed information is presented, the higher information value is assigned. The less lines or pages are used to present all the information, the higher presentation value is assigned. Once the information to be presented is determined, the presentation value can also be calculated. Therefore, for each possible combination of metadata to be presented, the total metadata value can be calculated using its information value and presentation value. That means, the best total metadata value $m(x, y)$ can be obtained by selecting the information combination that maximizes the total metadata value. $m(x, y)$ can be calculated as;

$$m(x,\ y) = \max\{\ f_m(V_I,\ V_P)\ \} \qquad \textbf{(13)}$$

where f_m is a function to determine the total metadata value considering the balance of the two factors for each possible combination of metadata.

We define the function f_m as follows;

$$f_m(V_I,\ V_P) = V_I \times V_P \qquad \textbf{(14)}$$

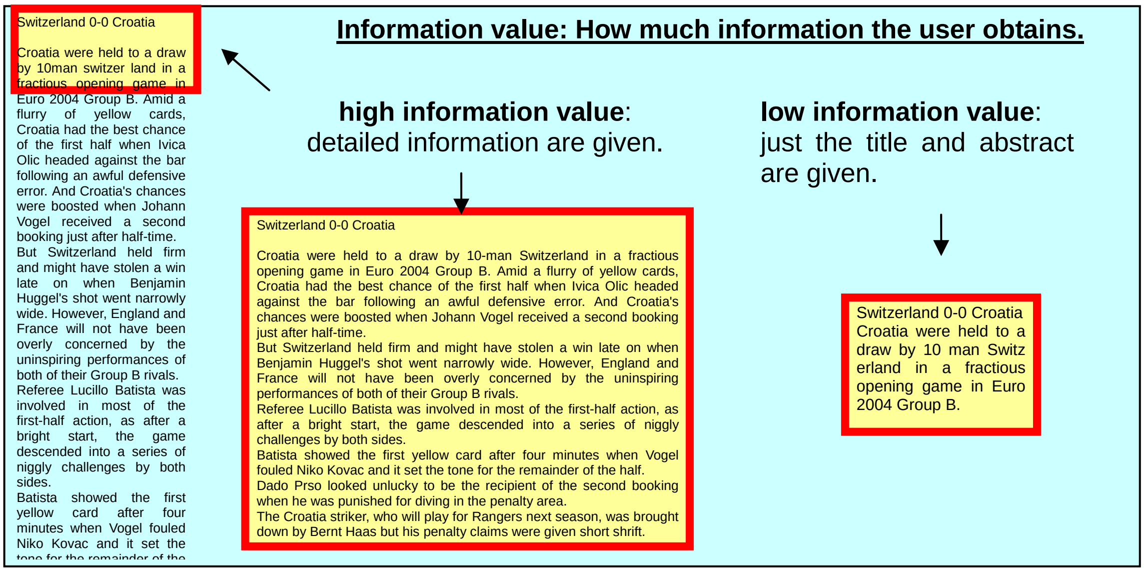


Figure 65. Example of information value

12.4.2. Information value

Information value represents the quantity and quality of information the user obtains. Figure 65 illustrates an example where information value is high and is low. When a large part of the metadata is presented, the information value becomes high. On the other hand, when just few information is given, the information value becomes low.

The information value also depends on the element to be presented. For example, a title would have high value (would be very important), a detailed text could have low value, and some specific data like URL could have higher value. Hence, the value for each element entity of the metadata needs to be defined beforehand. The information value is calculated by adding and normalizing the value of all the presented element entities. Information value V_I can be defined as follows;

$$V_I = \frac{\sum_{k=1}^{n} \{a_k \cdot v(k)\}}{\sum_{k=1}^{n} v(k)} \tag{15}$$

where

a_k: presence of element entity no.k. (normalized to 0-1) ,

$v(k)$: importance value of element no. k (normalized to 0-1) ,

n: number of elements in the metadata .

a_k is defined as how much the entity of the element k is presented. If all text of the element entity k is presented, a_k =1 is assigned while a_k =0 is assigned when this element entity is not presented at all. If some part of the element entity is presented, a_k is determined considering the amount of presence. For example, if only the first sentence out of all text of five sentences inside the element is presented, a_k could be defined as 0.2. The value $v(k)$ is assigned based on the importance of the element k. Taking a TV news content as an example, value for the topic title could be 1, the detailed text 0.5, and the URL leading to more related information 0.75.

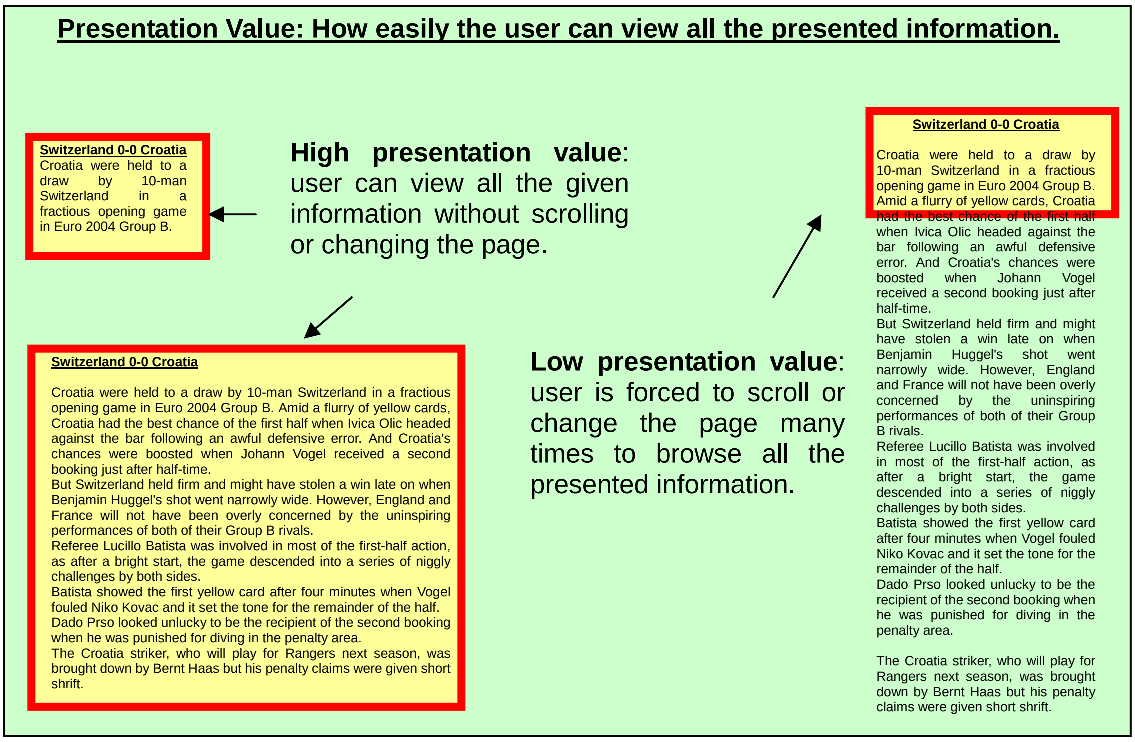


Figure 66. Example of presentation value

12.4.3. Presentation value

Presentation value represents how easily the user can view all the presented information. Figure 66 illustrates an example where presentation value becomes high or low. When a large part of the metadata is presented inside the screen, the presentation value becomes high. On the other hand, when just few information is shown inside the screen, the information value becomes low.

In metadata presentation, we consider how much screen space the presented text information requires. To be specific, we use the number of text lines necessary for presenting all the given information to evaluate the presentation value. The number of lines within the screen and also the number of characters within a line for displaying text information are necessary to calculate the necessary space for presenting the information. Presentation value V_P can be defined as a function of number of necessary text lines l. We define it as follows;

Presentation value V_P:

$$V_P = f_p(l)$$

$$f_p(l) = \begin{cases} 1 & (l < L) \\ (l_{\max} - (l - L))/l_{\max} & (L < l < l_{\max}) \\ 0 & (l > l_{\max}) \end{cases} \quad (16)$$

Parameters

l: Number of lines necessary for presenting all the text ,

l_{max}: Maximum number of lines allowed to be presented ,

L: Number of lines within the screen

C: Number of characters within a line .

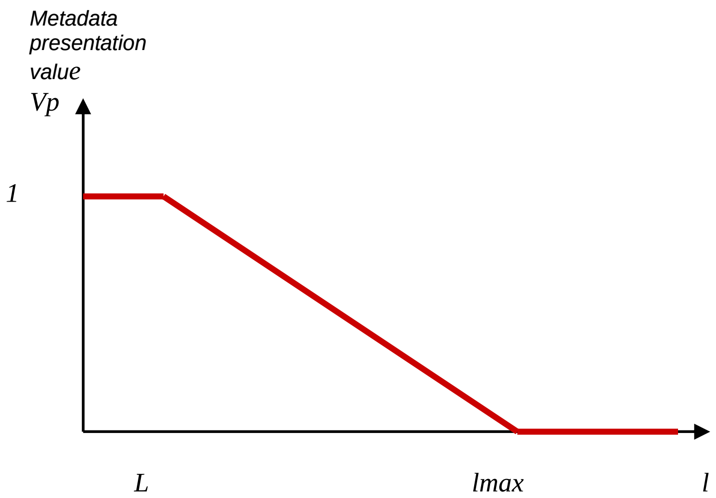


Figure 67. Relation between presentation value and necessary lines for presentation

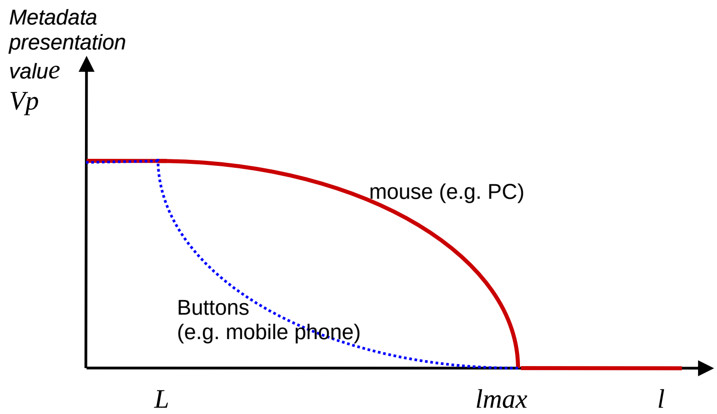


Figure 68. Example of relation among browsing interfaces, presentation value, necessary lines for presentation

Figure 67 shows the curve for function (16). The more lines are required to present the information, the less presentation value becomes. If the number of lines for presentation is larger than l_{max}, the presentation value becomes 0 as it is too difficult for the user to browse all the information.

In function (16) the presentation value is defined linear to the number of necessary lines, but it is obvious that it is not so simple the perceived presentation value for the user. Controlling the font size of the presented characters is an interesting extension. Controlling the presentation considering the hardware to interact with the browser is another interesting extension. Figure 68 illustrates some examples of the presentation value with different interfaces. With a mouse as an interface, it would not be so different for some users the presentation value to scroll 5 pages and 15 pages as the mouse enables easy scroll of the page. However, with buttons like in mobile phone, to scroll 15 pages would be much more than three times as tough as scrolling 5 pages as it is tough to scroll the screen. It is important to define an appropriate function f_p depending on the end-user device and input interfaces.

13. Evaluation

13.1. Prototype for performance evaluation

To evaluate the performance of our proposed scheme, we have developed a web-based prototype for browsing video content with its metadata. This prototype emulates any screen size given by the user. It consists of two parts; database entry of the contents, and adaptation server which includes the content adaptation process.

13.1.1. Database entry of the contents

Video contents and their corresponding metadata are prepared to be registered in the database. Each video file is parsed into items (scenes). We manually add a title and its detailed text description to each scene which forms a content metadata of each video file. The content metadata also includes a title and some description that represents the whole video content. One thumbnail that represents each video sequence is prepared for every sequence. Every video file is registered with their identification number, their content metadata and their thumbnails.

Figure 69 illustrates the structure of a content metadata of a video sequence. Each content metadata is analyzed to extract its number of layers, elements and characters. Note, in this implementation the number of layers is always 2; program layer and scene layer; however, it is obvious that more layers like group-of-shots, shots, key-frames, etc. are also applicable. Then we assign a value for each element in the metadata. The spatial resolution of the video content, the maximum/minimum allowed video scaling ratio, the number of layers and elements of the content metadata, the value and the number of characters of each element entity are registered as adaptation tools.

This prototype system manages all the content metadata and adaptation tools in a single relational database. As the content metadata and adaptation tools are highly structured, we use a relational database instead of a native XML database because it enables faster retrieval of the requested content and its description. All the XML descriptions of the content metadata described in MPEG-7 and adaptation tools described in MPEG-21 DIA are converted into RDB (MySQL) form for database entry.

Two tables in the database are prepared to describe the content metadata and adaptation tools of all registered video sequences. Table 3 describes the first table named *Sequence_data*. It contains all information about the video sequences. It contains the resolution, the text descriptions and all the adaptation tools of the video sequences.

Table 4 describes the second table named *item_data*. It manages the description about each item involved in the sequences. The title and text description for each item is stored with its value. The start frame number and its duration of the item within the video sequence are also stored to enable flexible video browsing of the desired item in the sequence.

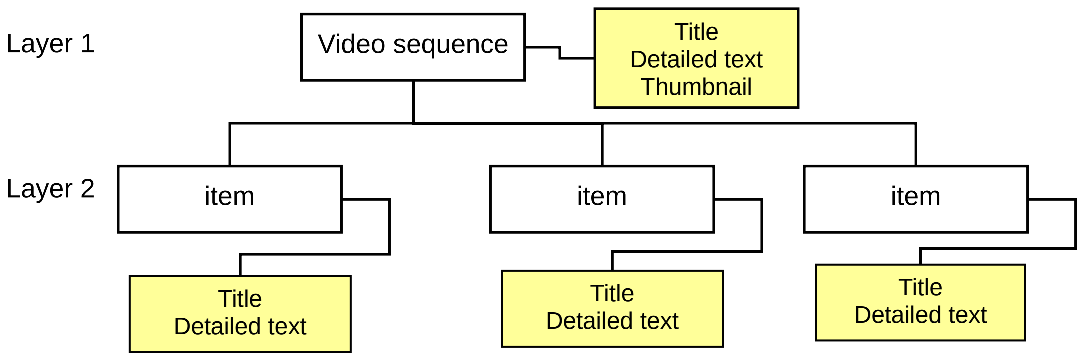


Figure 69. Example of content metadata used in the prototype

Table 3. This table manages descriptions of the video sequences (Sequence_data)

Column	Type	Semantics	
SequenceID	Integer	Unique ID key to identify video sequence.	PrimaryKey
SequenceURI	varchar(255)	Sequence URI. Used to locate the video file.	NOT NULL
MetadataURI	varchar(255)	Metadata URI. Used to locate metadata file.	
ThumbnailURI	varchar(255)	Thumbnail URI. Used to locate the representative thumbnail of the sequence.	
SequenceTitle	varchar(200)	Title of the video sequence	
SequenceTitleValue	float	Value of the title	
TextDescription	Text	Detailed text.	
TextDescriptionValue	float	Value of the detailed text.	
Width	Integer	Width of video frame.	NOT NULL
Height	Integer	Height of video frame	NOT NULL
MaxScale	Float	Maximum allowed scaling ratio of video	
MinScale	Float	Minimum allowed scaling ratio of video	NOT NULL
MaxPage	Integer	Maximum allowed number of pages for metadata presentation	

Table 4. This table manages descriptions of items in video sequences (Item_data)

Column	Type	Semantics	
SequenceID	Integer	Sequence ID which includes the item.	NOT NULL
ItemID	Integer	Unique ID to identify each item.	PrimaryKey
ItemTitle	varchar(200)	Title of each item.	
ItemTitleValue	Float	Value of the item title.	
TextDescription	Text	Detailed text of each item.	
TextDescriptionValue	Float	Value of the detailed text.	
StartFrameNumber	Integer	Start Frame Number.	
Duration	Integer	Duration of each item.	
NumOfChars	Integer	Number of characters of the detailed text	

Table 5. This table manages descriptions of shots in items or video sequences (shot_data)

Column	Type	Semantics	
SequenceID	Integer	Sequence ID which includes the shot.	NOT NULL
ItemID	Integer	Item ID which includes the shot.	
ShotID	Integer	Unique ID to identify each item.	PrimaryKey
ShotTitle	varchar(200)	Title of each shot.	
ShotTitleValue	Float	Value of the item shot.	
TextDescription	Text	Detailed text of each shot.	
TextDescriptionValue	Float	Value of the detailed text.	
StartFrameNumber	Integer	Start Frame Number.	
Duration	Integer	Duration of each shot.	
NumOfChars	Integer	Number of characters of the detailed text	

Table 6. This table manages descriptions of variations derived from the original video sequences (Variation_data)

Column	Type	Semantics	
SequenceID	Integer	Sequence ID of the original source.	
VariationID	Integer	Unique ID to identify each variation.	PrimaryKey
VariationURI	varchar(255)	Sequence URI. Used to locate the video sequence.	
Width	Integer	Width of video of the variation.	
Height	Integer	Height of video of the variation	
MaxScale	Float	Maximum allowed scaling ratio of video	
MinScale	Float	Minimum allowed scaling ratio of video	
VariationValue	Float	Value of the variation	

In case there are more layers in the metadata, a table with the same structure is prepared.

Table 4 can be added to manage the metadata of this layer. Table 5 shows an example when shots are described under items. In case there are variations of the video, we use Table 6 to manage the variations derived from the original video sequence. Each variation is linked with the original video sequence. The allowed scaling ratio and its variation value are also stored for supporting video content adaptation.

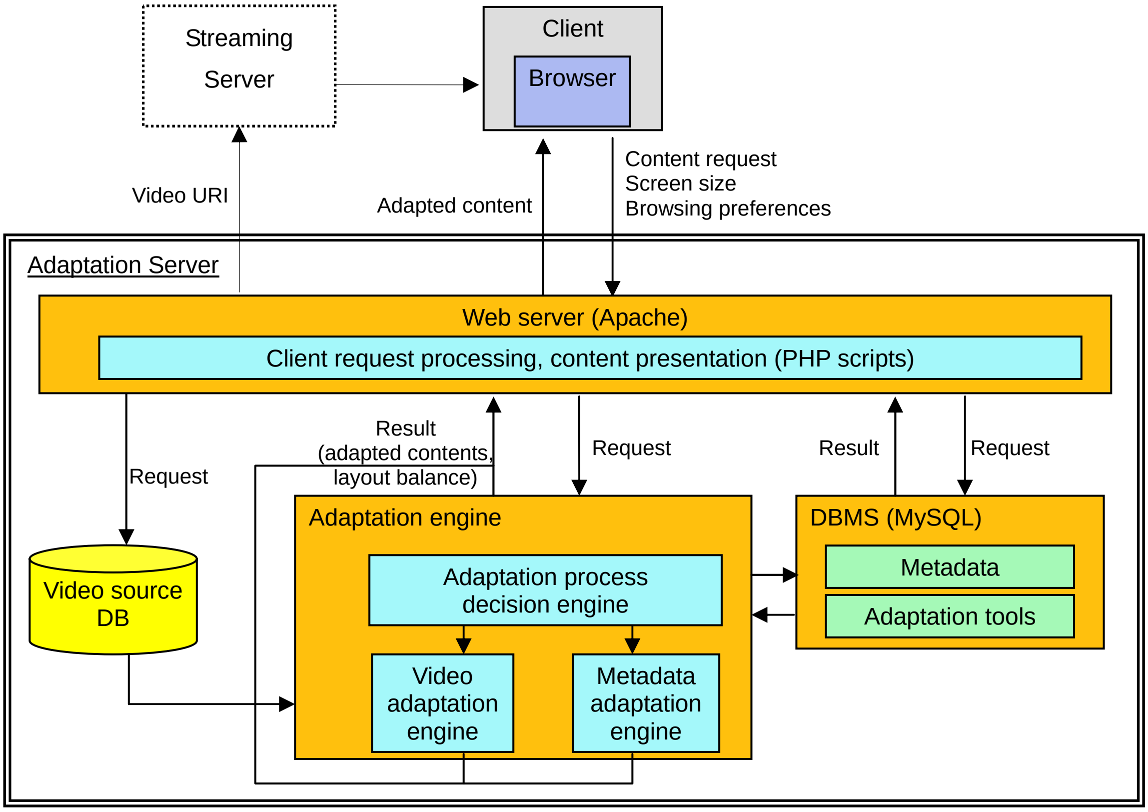


Figure 70. System architecture of the prototype adaptation server

13.1.2. Adaptation server

The adaptation server is implemented as a Web server. The adaptation server receives the id number of the requested content, the display capability (screen size) and the browsing preferences posted by the user. Then, the server dynamically scales the requested video content by the consumer, scales its content metadata, and balances the presentation of the video and its metadata as shown in Figure 42. The balance and scaling way is determined considering all the adaptation tools of the requested content and the received data from the user.

Our goal is to develop a system that automatically provides to any terminal of any user their desired video content with its metadata instead of giving them manually and post to the server. In this prototype, we evaluate the performance by giving manually the screen size and emulate the adaptation in a window of the given screen size. By accepting manually the screen size, it becomes possible to evaluate the adaptation performance towards any potential screen size. The terminal capabilities and browsing preferences will be described in CC/PP or MPEG-21 DIA and will be read automatically.

Figure 70 presents the system architecture of the developed adaptation server. We use Apache as a web server, MySQL as a database management system, and Real Helix Server version 9 as a video streaming server. Real Producer 10 is used to create variations of video contents. The adaptation engine is developed by using C++. PHP scripts have mainly been utilized to realize flexible user interfaces. The adaptation engine is called from PHP scripts to perform on-the-fly adaptation of the requested content.

Figure 71. Content selection menu and manual description of the screen size

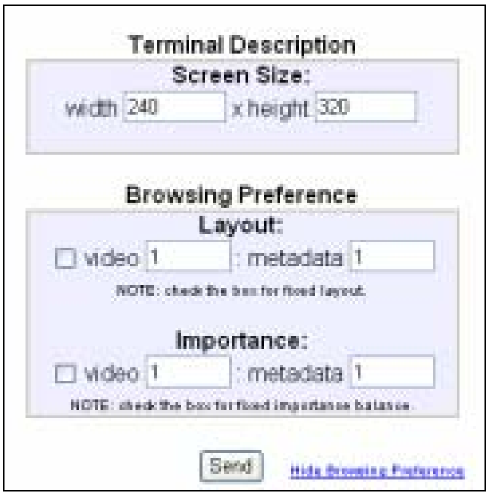


Figure 72. Editing form of browsing preferences. This form appears when "Edit Browsing Preferences" is clicked.

a) 

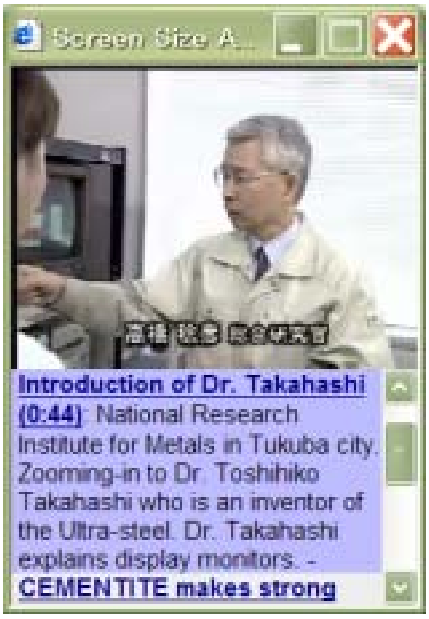
b) 

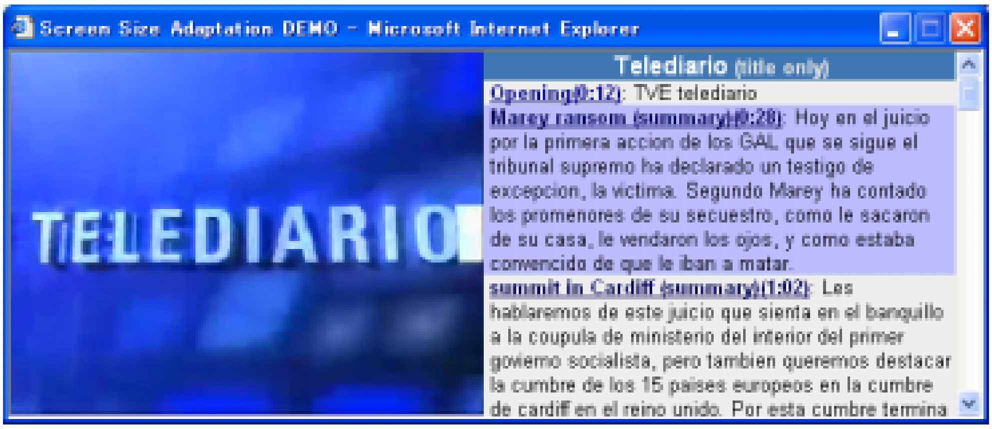
Figure 73. Dynamic presentation of the selected content under the given conditions. The system allows flexible video browsing of the interested topic by clicking its title. It also allows flexible metadata browsing by clicking the right "+" symbol of the interested topic.

Figure 71 shows a screenshot of the initial menu presented to the user when accessing the server. There are two forms in this menu; "Video content list" and "Terminal Description".

In "Video content list", the available contents called from the database are listed. Each content title, description and a representative thumbnail is displayed. This list could be created from a keyword search result, recommendation from the system, or just could be the available contents. The user has to select one of the listed content to browse. In "Terminal Description", the desired screen size needs to be given by the user in the form. If the user has some browsing preferences about the balance of video pane and metadata pane and of the importance of video and metadata, the user needs to click "Edit browsing preferences". The form in Figure 72 appears and the preferred values have to be filled in these forms.

After selecting one content and filling the screen size, and browsing preferences, if any, "SEND" button needs to be pressed. The filled values are sent to the web server with the id number of the selected content. The number of characters and number of lines of text that can be displayed are defined beforehand.

A pop-up window of the given size appears on the screen. The adaptation server adapts and displays the selected content in this window. Figure 73(a) shows a screenshot of the adapted content in a window of 240x320 pixels. Only the title for each item is presented in this example due to limitations of space. Each title presented in the metadata pane is linked with the corresponding part of the video content. The user can access the video of the desired item by clicking its title. The corresponding part of the video begins to be delivered by the streaming server. It is also possible to browse the detailed text of the interested topic by clicking the right "+" symbol as each title is also linked with its detailed text of the same item (Figure 73(b)). By clicking the "-" symbol in the detailed text, this detailed text is hidden.

i) 

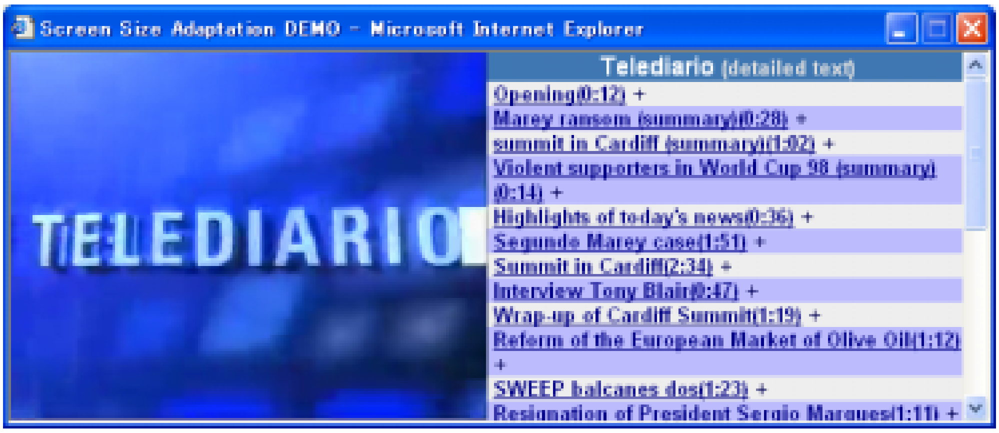
ii) 

Figure 74 i) Dynamic presentation of the selected content under the given conditions (640x240). All the titles and their detailed text are presented. ii) The display mode of the metadata can be switched by clicking “title only” and “detailed text”.

Figure 74 shows a screen shot of the same content in a window of 640x240 pixels. All the titles and their detailed text are presented as there is enough space to present all of them. By clicking “title only” in Figure 74 (i), the display mode of the metadata can be switched. As shown in Figure 74 (ii), just the titles of the items are presented in this mode. It is also possible to browse the detailed text of just the interested topic by clicking the right “+” symbol. By clicking “detailed text” in Figure 74 (ii), the display mode of the metadata can be switched again to Figure 74 (i).

Thus, flexible browsing of a video content and its metadata is achieved.

Table 7. Relation between presented metadata and its information value

Presented metadata	Metadata information value
No information	0
Program title	$1 / 2(n+1)$
Program title + program details	$2 / 2(n+1) = 1 / (n+1)$
Item title	$n / 2(n+1)$
Program title + item title	$(n+1) / 2(n+1) = 0.5$
Item title + item details	$2n / 2(n+1) = n / (n+1)$
Program title, details + item title, details	1

Table 8. Metadata parameters

Title element value	$v=1$
Detailed text element value	$v=1$
Number of items in a program	n
Maximum lines (pages)	$l_{max} = 30\ L\ (30\ pages)$

Table 9. Video parameters

Maximum scaling ration	*2*
Minimum scaling ration	*0.5 or 0.25*
Variations	*none*

13.2. Adaptation performance evaluation

We evaluate the adaptation performance by adapting various contents to various screen sizes. To evaluate the influence of the video content resolution to the adaptation result, we prepare video contents with various resolutions. To be specific, the resolutions of the prepared video content are 176x144, 320x240, 352x240, 352x288, 640x480, 704x480 pixels/frame. Note, 176x144 and 352x288, 352x240 and 704x480, 320x240 and 640x480 have the same aspect ratio. The maximum scaling ratio is set to 2 and the minimum ratio to 0.5 for or 0.25.

For metadata, four types of metadata, program title, program detailed text, item title, item detailed text are prepared. We assign the same value to all of them. Let n be the number of items involved in a certain program. There are $2n$ item elements and 2 program elements that can be presented. We select the metadata to be presented from seven possibilities given in Table 7 considering the tradeoff among the metadata pane size, information value and the total number of lines required for presentation. The metadata information value corresponding to each possibility is also presented in Table 7.

These video contents and metadata are adapted to various screen sizes; 240x320 (latest mobile phone), 320x240, 640x240 (Handheld PC), 240x640, 640x480 (VGA), 480x640, 720x480 (TV), 480x720, 800x600 (SVGA), 600x800 (PC browser), 1024x768 (XGA, PC screen). Figure 75 and Figure 76 shows some adaptation results of a single content to various screen sizes.

Figure 75. Adapting the same video content and metadata to various screen sizes of a) 240x320, b) 320x240, c) 640x240, d) 240x640, e) 480x640

Figure 76. Adapting the same video content and metadata to various screen sizes of f) 640x480, g) 720x480, h) 600x800, i) 800x600, j) 1024x768

Figure 75 and Figure 76 shows the results of adapting the same video content (352x288) and metadata to these screen sizes.

For video content presentation, video content size is maximized to fit the screen in size (a) 240x320, (c) 640x240, (d) 240x640, (e) 480x640, (h) 600x800. The same results as modality (video) centric approach can be obtained in these cases. For screen size of (b) 320x240, (f) 640x480, (g) 720x480, (i) 800x600, the video content size is balanced with metadata to maximize the total content value. As “2” is assigned as the maximum scaling ratio, the maximum video content resolution is 704x480. If the screen size becomes larger than a certain amount, the video content resolution is always transformed to 704x480 and the rest of the screen is used to present the metadata. Therefore, in screen size (j) 1024x768, the video content is transformed to 704x480 and the rest is used for metadata. If the original video content size is for example, 704x480, then different results are expected as the video content can be transformed up to 1408x960. In this prototype the content is sent to the client as is, and the video size is resized in the client’s browser.

For metadata presentation, only titles of each item are presented in small screens like (a) 240x320 and (b) 320x240. For the other larger screen sizes, all information is presented as all information can be presented within 15 pages in those screen sizes. This capability to present all the information within 15 pages make the metadata presentation value larger than 0.5. Even if all the titles can be presented in a screen, the total metadata value becomes 0.5 for the metadata prepared for this content. It quite depends on the amount of the prepared metadata and on the value of maximum pages. If there are a large number of items involved in the content, only titles would be presented even the screen size becomes larger. If there are just a few items, all text would be presented in even small screens. Note, the font size of the metadata pane is enlarged in case the necessary pages for presentation are lower than a certain threshold (see h,i and j in Figure 76). In screen size (b) 320x240, (f) 640x480, (g) 720x480 and (i) 800x600, the minimum metadata pane size where metadata information value is maximized (information value=1) are selected as the best balance between video content and metadata presentation.

Considering those results, in most cases the best balance between video and metadata presentation are obtained when the video content size is maximized or when metadata information value is maximized. This result is useful for accelerating the determination process of p, the layout balance between video and metadata pane. Instead of calculating the total content value for every possible p, we can first try the following two approaches and select the better one to reduce computational cost;

1. Maximize the video content size within the screen (video-centric approach), use the rest of the screen for metadata presentation, and then calculate the total content value.
2. Calculate the minimum pane size where metadata information value is maximized, use the rest of the screen for video presentation, and then calculate the total content value.

Another solution for accelerating this process is as follows;

1. Calculate the lower value p $(=p1)$ that makes metadata presentation value higher than 0.
2. Calculate the higher value p $(=p2)$ that makes metadata information value higher than 0.
3. Calculate the lower value p $(=p3)$ that makes the video content value higher than 0.
4. Calculate the higher value p $(=p4)$ that makes the video content value higher than 0.
5. Calculating total content value for p that meets “$max(p1,p3) < p < min(p2,p4)$” instead of calculating them for “$0<p<1$”.

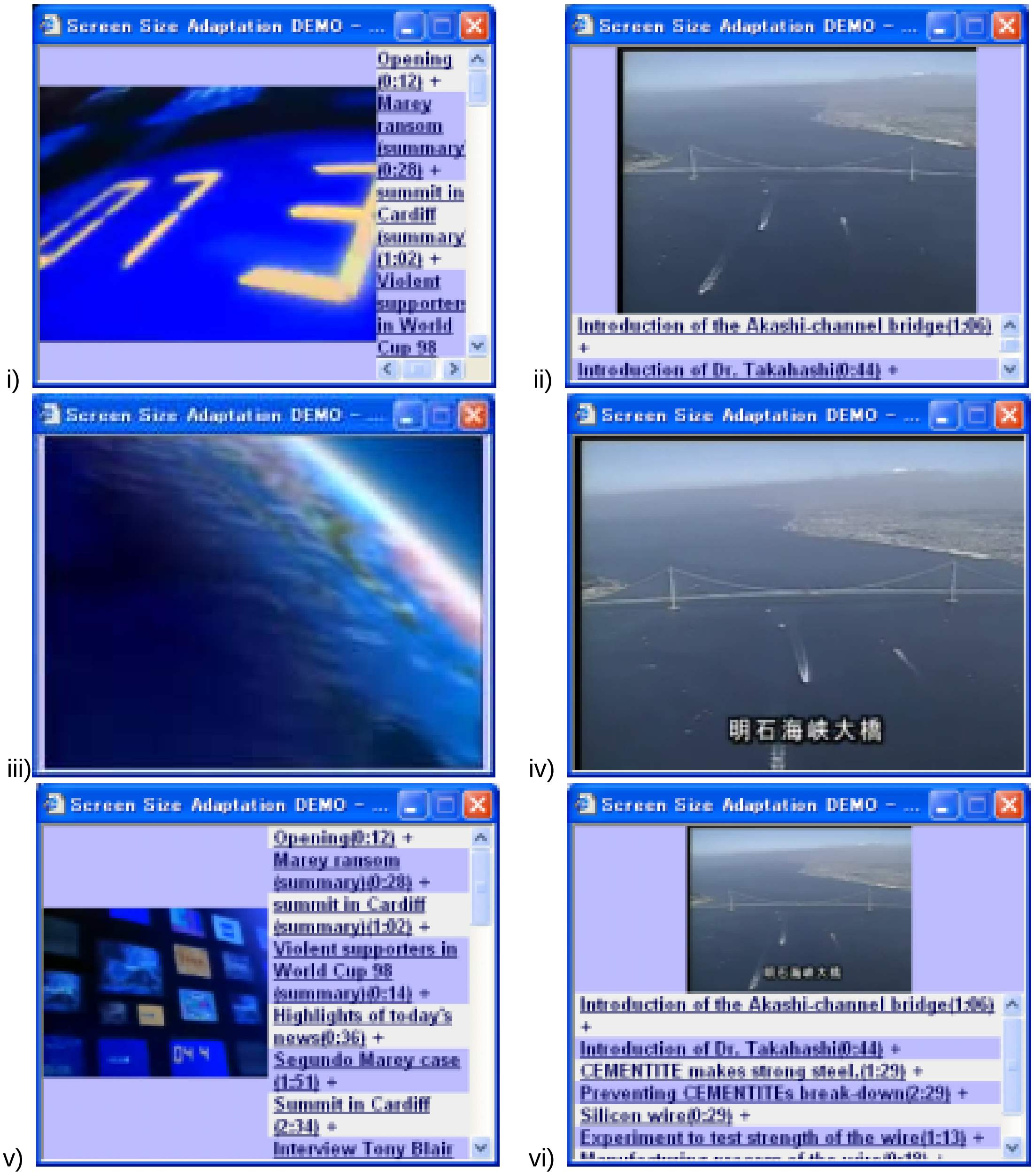


Figure 77. Adapting different contents to screen of 320x240 pixels. i) video content resolution: 352x288 pixel/frame, ii) video content resolution: 320x240 pixel/frame, iii)-iv) video-centric approach, v)-vi) layout-centric approach (1:1)

Figure 77 shows some results of adapting different contents to a fixed screen size of 320x240 pixels. In the screenshot i) and ii) the resolution of the original video content differs. The former one is 352x288 and the latter 320x240 pixel/frame. Due to the difference of the aspect ratio of the original content, different layout is selected to present the video content and its metadata. Iii) and iv) illustrates an adaptation result when video-centric approach is taken. Only the video content is displayed. v) and vi) presents when a user wants a fixed balance of video and metadata presentation (1:1). Layout-centric approach is utilized in this case.

13.3. Computational cost evaluation

We evaluated the computational cost of the adaptation process. Table 11 shows the average processing time of adaptation to various screen sizes using an Intel centrino 1.5 GHz machine as an adaptation server. This processing time is calculated by performing 20 times the adaptation to the same screen size and then averaging its processing time. All possibilities in dividing the screen both horizontally and vertically are processed to determine the best layout. That means, for screen size (X,Y), there are $X+Y$ possibilities of *p* that are used to calculate the total content value. The metadata used for evaluation consists of 22 news items with title and detailed text (Table 10). Figure 78 and Table 11 shows the obtained results by adaptation to various screen sizes. Figure 78 denotes that the processing time increases linearly to the screen size. Furthermore, it indicates that the adaptation time is less than 1 second even in adaptation to huge screens like 2048x2048 pixels. This result seems quite positive considering the response time the user allows to wait after requesting the content in web application is about 5-10 seconds.

Table 10. Number of characters in each item in the metadata used for evaluation

News Item No.	Title	Detailed text
1	24	598
2	31	437
3	25	441
4	17	265
5	22	391
6	14	368
7	66	403
8	50	301
9	12	431
10	12	427
11	15	419
12	18	108
13	79	409
14	12	416
15	49	438
16	12	385
17	20	336
18	21	330
19	9	276
20	17	264
21	12	194
22	20	447

Table 11. Processing time of adaptation to various screen sizes

Screen size	Possible *p*	Adaptation time
144x176	320	89 msec
240x320	560	157 msec
640x240	880	250 msec
480x640	1120	290 msec
640x480	1120	290 msec
720x480	1200	299 msec
800x600	1400	319 msec
1024x768	1792	363 msec
1600x1200	2800	476 msec
2048x2048	4096	617 msec

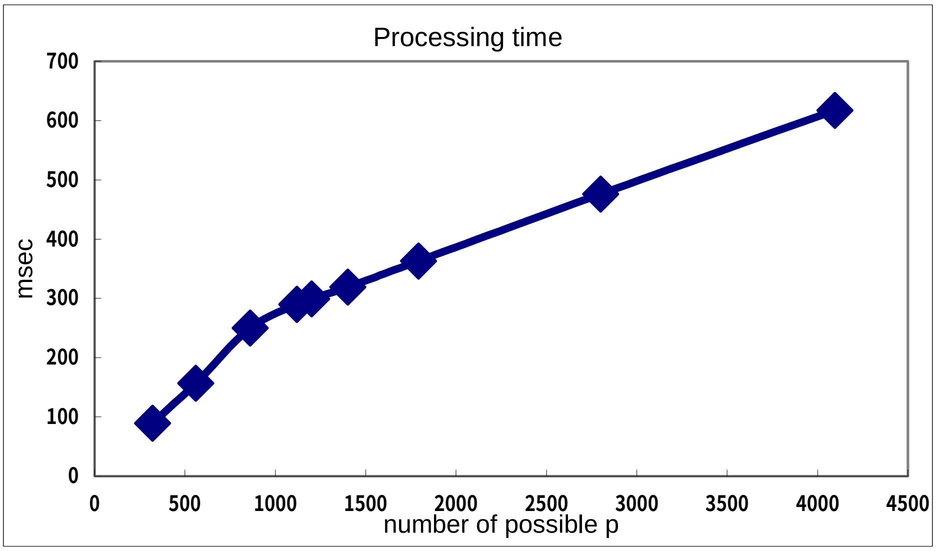


Figure 78. Processing time of adaptation to various screen sizes.

13.4. Discussion

The quality metric initially used in this experiment was based on a “value” which compares the adapted content (video, metadata) with the original content. When the original video content is presented as is, and all the metadata value is presented within a screen, the total content value 1 is assigned.

This metric works when the resolution of all the prepared video content and the amount of metadata are distributed in a small range. TV news archive is a good application because the video content materials are prepared with same resolution in most cases. Also, as the length of each news program is relatively the same, the amount of metadata does not vary so much. However, this metric causes problems in the following cases;

1. When the size of the original video content varies quite a lot.
2. When the size of the metadata varies quite a lot.
3. When small video content has to be displayed in large screen.

A typical case for 1 and 2 would be as follows; One case the original video content is 704x480, and it is adapted in a screen size of 176x144. The other case the content is 352x288 and the screen size is the same, 176x144. Unless the perceived quality of the adapted video would be more or less the same in both cases, the content value of the former case would be the half of the latter case. In this case, a quality metric that evaluates the absolute content value without references is necessary. The same thing happens in metadata.

About 3, the definition of content value of upscaled video contents plays a key role. If the value is increased in proportion to the adapted video size, video would be too dominant in terms of total content value. If the value of the original size is defined as maximum and the value remains the same in case of upscaling, then the video content would never be upscaled and metadata pane would be too dominant. Currently, we normalized the video content value using the maximum displayable scaling ratio.

Both evaluation metrics with reference and without reference should have to be taken into account considering the nature of the data and the target application.

14. Conclusion

We have presented a framework for joint adaptation of an audiovisual content and its metadata. The adaptation of the audiovisual content and its metadata are balanced to fit the given screen size in a way that maximizes user experience in browsing the desired content. The desired audiovisual content and its metadata are adapted using three types of descriptions; content description, adaptation tools to support the adaptation process, and description of the consumer side which includes device screen size and browsing preferences. The adaptation process is modeled as an optimization problem of the total value of the content provided to the user. The total content value is maximized by jointly controlling the balance between video and metadata presentation, the adaptation way of the video content, and the quantity and quality of metadata to be presented considering the device screen size and browsing preferences. Experimental results show that this scheme enables users to browse audiovisual contents with their metadata optimized to the screen size of their devices.

Future works includes quality assessment methods, extension of adaptation methods, and extension of modalities and constraints. In quality assessment, subjective experiments to determine the metric define the value of the perceived content are emerging. Video content adaptation considering ROI and visual attention model are another interesting direction instead of just scaling the video content. Extension to more complex layouts for presenting multimedia contents and the adaptation to any network (bandwidth) are also important from the practical point of view.

Part III: Universal Multimedia Access Applications

15. Introduction

This part introduces some potential UMA applications. As multimedia archives already have a rich set of metadata, these archives are expected to be extended to universal access services with minimum cost. We choose universal access applications to TV news archives as an example to make clear where the added value in UMA application exists.

16. Universal Access application to TV news archives

16.1. Current TV news delivery services

More and more multimedia news are distributed to various terminals. A lot of broadcast stations provide their news contents also on the web (e.g. BBC, CNN, TSR, etc.). Thanks to the advancement of network infrastructure, TV news delivery services not only to PC but also to portable devices like mobile phones are also beginning to be considered.

Figure 79 illustrates a typical TV news delivery web site. In the front page, a list of news (latest news, categorized news (sport, politics, …), and a keyword search engine so that the user can search his/her interested topic is prepared. After one of them being selected, the news video is presented with some text annotations. The big difference with analog television is the interactivity. The user can access and browse the desired part of his/her interested topic.

Figure 80 describes about metadata creation and the added value which these metadata can provide in these news delivery services. The following types of metadata are considered to be used in these services;

A. Structural information:

A news video program is a well-structured content. Each program is divided into openings, summary, news topics, and so on. This structure is described for each news program.

B. Creation information:

Creation information like the duration of each topic, the whole video, the created date and time are described for two purposes. The first purpose is for content management and search purpose. The second is to provide the user the length of each news topic so that the user can estimate its duration.

C. Text annotation for each topic:

Text annotation is also used to enable topic search and to provide additional information of the topic to the user.

D. Category of each topic:

The category is added so that the user can access the desired category news.

With these metadata, the following services are provided as added value in news archives;

- TV News with topic list browsing.
- Date/time based news browsing. (19:30 news of 29/02/2004)
- Keyword search based browsing.
- Category based browsing. (Swiss, World, Economy, Sports, Culture, Regional)
- Latest news browsing.

Figure 79. Typical TV news delivery web site (TSR.ch)

Figure 80. Metadata creation and added value in current news delivery applications

However, under the current situation, the contents are designed separately for each device. This situation let the news provider make more and more different versions of the same news in proportion to the target devices. As a rich set of metadata is already assigned to each news contents, these metadata could be quite useful to extend the conventional news delivery systems to universal access systems.

16.2. How to maximize user satisfaction in news delivery applications

To maximize user satisfaction in news delivery applications, the satisfaction of both user and provider has to be considered. For the user side, the main factors of satisfaction exist in how easily the user can access the **desired** news contents **anytime, anywhere, from any device**. For the provider side, it is important to make the **additional manual cost** as **minimum** as possible while maximizing user satisfaction.

16.3. New added value on TV news archives

To maximize user satisfaction in news delivery applications, it is quite important to add new values to multimedia news archives with minimum additional manual cost. We have considered the following five added values. The first one is to enable Universal Access to multimedia news archives with minimum additional manual cost. The rest are user assist service to increase the accessibility to the desired information.

I. Universal Access service (Terminal and network independent access).

   Universal access service means the capability to access news archives from any terminal under any network conditions.

II. Flexible news browsing and navigation service.

   Flexible news browsing enables browsing the desired content independent of time and space scale, from one's preferred viewpoint, and to browse as much as one wants.

III. Personalized news delivery service.

   Personalized news delivery service enables to send personalized news contents to each person. Personalization includes adapting to user preferences and presentation preferences, adapt difficulties in hearing and seeing, news recommendation considering their usage history, and so on.

IV. Natural environment based news delivery service.

   Natural environment based news delivery service enables to send news contents based on their natural environment. For example, location based news delivery service, (exact position based service (movie theater, shops, monuments…), regional information, and country level (JPN, US, Switzerland, etc)), current time based news delivery service, and so on.

V. Mobility service.

   Mobility service enables to continue watching the interested news contents from the current state using different devices. (e.g. watch the news in PC at workplace -> continue watching the same content with mobile phone in the train -> watch the rest in TV at home.)

17. Other examples of Universal Multimedia Access applications

During the investigation on UMA, several applications considering the difficulties in the author's first days in Lausanne, concerning language problem and understanding Swiss culture, were discussed ("whom" = "myself").

The first one is a system that provides universal multimedia e-learning experience that enables to study French courses everywhere with my pace. The second one is an information service for new-comers (e.g. TV contents selection for Lausanne new-comers). The third one is on creation and delivery of barrier-free content for handicap support. Handicap does include not only audiovisual impairments, but also difficulties on understanding language and cultural background.

17.1. Universal multimedia e-learning experiences

The first application is a system that provides universal multimedia e-learning experience that enables to study French courses everywhere with one's pace. This system can be generalized not only for French studies, but to any multimedia e-learning contents.

Table 12 indicates the requirements for the universal multimedia e-learning system from the viewpoint of both user and service provider.

Table 12. User and system requirements for universal multimedia e-learning system.

User requirements	System Requirements
View the contents from any device at any place.	Adapt contents based on terminal capabilities and natural environments.
Receive the appropriate contents in terms of their semantics.	Select the appropriate content for each user based on user and content preferences.
Access to any part of the content.	Allow access to any part of the content and its annotation.
Continue access from where he/she was from any terminal.	Manage study progress. Store usage history. Session mobility.
Select the preferable modality (audio/video/text /graphics/etc.) and preferable playing speed.	Prepare or create preferable modalities for the user.

Technical requirements for this application are as follows;

- Content adaptation based on terminal capabilities and natural environments.
- Content selection based on user preference and usage history.
- Adaptive content and metadata visualization based on user's request to facilitate access.

- Transfer the state of Digital Items from one terminal to another. (session mobility)
- Preparation of multi-modal variation (video, audio, text) of the e-learning content or on-the-fly transmoding technology.
- Control of play speed.

17.2. Category profile (information delivery service to people of some category

When people come first at some new place, it is always difficult to find important information to set up their life and understanding the culture. This problem can be generalized as an information delivery service for people of a certain category. The main thing is to deliver the appropriate information for each people in each category, and also updating the profile of both personal and their category based on their experience. Table 13 indicates the requirements for the system from the viewpoint of both user and service provider.

Table 13. User and system requirements for category preference system.

User requirements	System Requirements
Obtain essential information for new-comers.	Creation of general category preference. (statistical? rule-base?)
Receive local information.	Content selection based on position and time(?).
Receive the appropriate contents according to the user's experience	Update of category profile and user profile based on the user's experience.

17.3. Barrier-free content for handicap support

Handicap here includes not only A/V impairments but also lack of ability to understand foreign languages and cultures. It is strongly desired that the contents are transformed in an understandable form for each user.

18. Conclusion

In this part, we have presented some potential Universal Multimedia Access applications. Taking TV news archives as an example, we analyzed the current added values in multimedia content archive systems and identified the new potential applications to these archives. Five applications have been identified to provide added value to multimedia archive systems; Universal Access service (Terminal and network independent access), flexible news browsing and navigation service, Personalized news delivery service, natural environment based news delivery service, Mobility service. Other examples of UMA were also presented in this part. Future work includes to find the best tradeoff between the additional cost and the quality of service that can be provided to the user.

References

References on adaptation

[Mohan99]
R. Mohan, J. Smith, C.S.Li, **“Adapting Multimedia Internet Content for Universal Access**”, IEEE Transactions on Multimedia, pp. 104-114, March 1999.

This paper present a system that adapts multimedia Web documents to optimally match the capabilities of the client device requesting them. This system has two key components, which are a representation scheme called the ‘InfoPyramid’ that provides a multimodal, multiresolution representation hierarchy for multimedia, and a customize engine that selects the best content representation to meet the client capabilities while delivering the most value.

[Lee01]
Keansub Lee; Hyun Sung Chang; Seong Soo Chun; Hyungseok Choi; Sanghoon Sull; “**Perception-based image transcoding for universal multimedia access”,** Image Processing, 2001. Proceedings. 2001 International Conference on , Volume: 2 Page(s): 475 -478 vol.2, 7-10 Oct. 2001.

Abstract:
This paper proposes an image resizing method which fits the client display size considering the object importance inside the image. Importance values are added beforehand to each block, and the resizing process is performed by combination of cropping and scaling using the proportion that maximizes the content importance values. (same idea as Region-of-Interest(ROI) in JPEG2000.)

Contents:
- perceptual hint for image transcoding
  - spatial resolution reduction value
  - transcoding hint for each image block
- Image transcoding algorithm based on perceptual hint
  - content value function
  - content adaptation algorithm
- experimental results

[Bruijin02]
O. de Bruijin, R. Spence, M.Y.Chong: “**RSVP Browser: Web Browsing on Small Screen Devices**”. ACM Personal and Ubiquitous Computing,Vol.6, Issue 4, pp. 245 - 252, 2002.

This paper illustrates the use of space-time trade-offs for information presentation on small screens using Rapid Serial Visual Presentation (RSVP) for Web browsing. The principle of RSVP browsing is to present objects in the image in an optimal order and presenting time.

[Fan03]
Xin Fan, Xing Xie, Wei-Ying Ma, Hong-Jiang Zhang, He-Qin Zhou: “**Visual Attention Based Image Browsing on Mobile Devices**”. IEEE ICME 2003, Vol.1, pp. 53-56, 2003.

This paper proposes a method for browsing a large image on small screen by extracting important regions/objects from the image, which and shown in a slide-show way considering their importance. The importance is determined considering saliency, face and text inside the regions.

[Liu03]
Hao Liu, Xing Xie, Wei-Ying Ma, HongJiang Zhang: “**Automatic browsing of large pictures on mobile devices**”. ACM Multimedia 2003: pp.148-155, 2003.

This paper proposes an image scaling method and how to optimize the order of browsing objects or parts of large images on small displays. The image is cropped to extract the significant parts of the images and then the cropped part is down-scaled. The balance between cropping and down-scaling is optimized to provide as much information as possible under resource constraints (JPEG2000 ROI-like). As an image browsing method, the optimal order and time of presenting image objects are studied based on image attention model (eg. Face and text, saliency objects presented first).

[Cavallaro03]
Andrea Cavallaro, Olivier Steiger and Touradj Ebrahimi, "**Semantic Segmentation and Description for Video Transcoding**" Proc. of the IEEE Int. Conf. on Multimedia and Expo, ICME'03, vol. 3, pp. 597-600, Baltimore, USA, July 6-9, 2003.

This paper summarizes video converting methods for adaptation to various bandwidths and terminal characteristics. Three methods are introduced; 1) signal-based conversion: spatial conversion, temporal conversion, color depth reduction, 2) object-based conversion: Control of coding conditions or decoding order of objects or regions of interest inside the image/video, 3) description-based conversion: Delivering/using just the features extracted from the objects (object identifier and shape information given as an example).

[ChenICME02]
Chen, R.Y.; van der Schaar, M.; “**Resource-driven MPEG-4 FGS for universal multimedia access**”, Multimedia and Expo, 2002. ICME '02. Proceedings. 2002 IEEE International Conference on , Volume: 1 , 26-29 Aug. 2002, Page(s): 421 -424 vol.1

Abstact:
This paper presents a method how to control the decoding complexity level which adapts the constraints of device capabilities. The selection of the best tradeoff among rate, distortion and complexity space based on constraints are necessary. The computational complex functions in decoding process consists of two categories, 1) bit-oriented (bit-fetch, VLD, bit-plane assembly)

proportional to transmission bit-rate, 2) frame-oriented (IDCT, reconstruction) proportional to the size of VOP(frame). For frame-oriented adaptation, different IDCT and reconstruction algorithms are selected based on complexity levels. Experimental results show that the decoding complexity can be accurately predicted based solely on the transmission rate even the video characteristics are quite different. Given a certain receiver capability and a desired decoding complexity level, the proposed system fits the receiver complexity constraints in addition to the network bandwidth.

[ChenICCE02]
Chen, R.Y.; van der Schaar, M.; "**Complexity-scalable MPEG-4 FGS streaming for UMA",** Consumer Electronics, 2002. ICCE. 2002 Digest of Technical Papers. International Conference on , Page(s): 270 -271, 18-20 June 2002.

Abstract:
This paper presents a method how to control the decoding computational complexity level which adapts the constraints of device capabilities. The computational complex functions in decoding process consists of two categories, 1) bit-oriented (bit-fetch, VLD, bit-plane assembly) broportional to transmission bit-rate, 2) frame-oriented (IDCT, reconstruction) proportional to the size of VOP(frame). For frame-oriented adaptation, different IDCT and recounstruction algorithms are selected based on complexity levels. Experimental results show that the decoding complexity can be accurately predicted based solely on the transmission rate even the video characteristics are quite different. Given a certain receiver capability and a desired decoding complexity level, the proposed system fits the receiver complexity constraints in addition to the network bandwith.

[Bjork00]
Niklas Bjork and Charilaos Christopoulos, "**Video Transcoding for universal multimedia access**", ACM Multimedia 2000.

This paper discusses the issue of adapting video streams to different type of terminals with different terminal capabilities such as screen size, amount of available memory, processing power and type of network access. Rate reduction model and resolution reduction model for transcoding are examined.

[Ferman02]
A. Mufit Ferman, James H. Errico, Peter van Beek, M. Ibrahim Sezan, "**Content-based Filtering and Personalization Using Structured Metadata**", JCDL'02.

This paper presents a framework for multimedia content personalization. This system consists of a profiling agent, which determines a user's profile from his/her content usage history, and a filtering agent, which filters contents according to the user's profile.

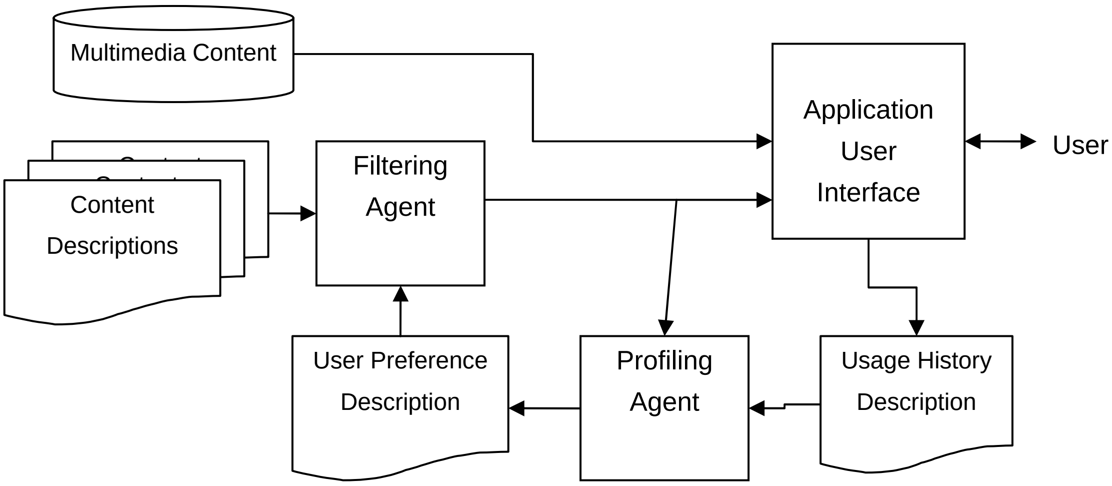


Figure A1. Content personalization framework

[Ferman03]
A. Mufit Ferman, Peter van Beek, James H. Errico, M. Ibrahim Sezan, "**Multimedia Content Recommendation Engine with Automatic Inference of User Preferences**", ICIP 2003, TP-P8.

This paper proposes algorithms for automatically determining a user's profile from his/her content usage history (profiling agent) and for automatically filtering content according to the user's profile (filtering agent). The profiling agent calculates the preference value for each program, for each kind of descriptor and for each category using the way and the amount of time the user interacts with given content. The profiling agent can also update the user preference dynamically, by considering the usage history items logged since the last update. The filtering agent compares each component of a user preference description with that of a content description, combines the individual test results into a single score that reflects the degree to which content fits the user's overall preference.

[Homayounfar03]
Homayounfar, K.; "**Rate adaptive speech coding for universal multimedia access",** Signal Processing Magazine, IEEE , Volume: 20 Issue: 2, Page(s): 30 -39, March 2003.

Abstract:
This paper is focusing on rate adaptation for mobile network.

- Transcoding (real-time, non real-time)
- Adaptive Multirate codec (AMR) adaptation.
- Content-based adaptation (user preferable codec conditions).

[Lienhart04]
R. Lienhart and I. Kozintsev, Intel Corporation: **"Self-aware Distributed AV Sensor and Actuator Networks for Improved Media Adaptation"**, ICME 2004 special session on Content Understanding and Transcoding Techniques for Media Adaptation.

This paper tries to combine multiple I/O devices into a distributed array of sensors and actuators and to shift the media adaptation research away from a single device/stream paradigm towards array multimedia processing. The way how to transform a network of off-the-shelf devices into a distributed I/O array by providing common time (with tens of microseconds precision) and 3D space coordinates (with a few centimeters precision) is presented.

[Shaar02]
van der Schaar, M.; Radha, H.; **“Adaptive motion-compensation fine-granular-scalability (AMC-FGS) for wireless video”,** Circuits and Systems for Video Technology, IEEE Transactions on , Volume: 12 Issue: 6 , Page(s): 360 -371, June 2002

Abstract:
This paper tries to introduce MC(motion convensation) within MPEG-4 FGS.
With two-loop MC-FGS the image quality becomes better about 2db but the decoding complexity increases about two times more.

Contents:
- MC-FGS structure
  - two-loop MC-FGS for B frames
  - single loop MC-FGS
  - Adaptive MC-FCS

[SumISCAS03]
Sun, H.; Vetro, A.; Asai, K.; **"Resource adaptation based on MPEG-21 usage environment descriptions",** Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on , Volume: 2 , Page(s): II-536 -II-539 vol.2, 25-28 May 2003

Abstract:
This paper addresses key issues for "Video Transcoding" for resource adaptation, with the background of UMA concept and MPEG-21 DIA. The major problem for UMA is to fix the mismatch between the content formats, the conditions of transmission networks and the capability of receiving terminals. MPEG-21 DIA aims at fixing these gaps by providing the standardized descriptions and tools for resource adaptation and descriptor adaptation. The key design goals of transcoding include two aspects, 1) to maintain the video quality during the transcoding process, and 2) to keep complexity as low as possible. Technologies to achieve the best perceptual quality for any User besides the existing transcoding technologies which controls bit-rate, frame-rate and spatial resolution are introduced as "Transcoding QoS". The key issue for Transcoding QoS is to optimize parameters based on objective quality measures and/or user preferences, and transcoding of multiple streams.

Contents:
- Overview of UMA
- MPEG-21 DIA
- Resource Adaptation Engine
  - Transcoding background

- Transcoding QoS

[Vetro01]
Vetro, A.; Huifang Sun; "**Media conversions to support mobile users",** Electrical and Computer Engineering, 2001. Canadian Conference on , Volume: 1 Page(s): 607 -612 vol.1, 13-16 May 2001.

Abstract:
This paper provides application services and media conversion techniques for content delivery to mobile users.
Two types of services for mobile users are introduced, 1)content search and retrieval, and 2)push and filtering service. For search and retrieval, only the relevant parts of the content should be delivered with a playable format (ex. MPEG-4, text-based message, closed caption information). For push and filtering, the contents may be adapted differently depending on the user environment, which may include the location of the user and currently enabled device.
Media conversion needed to support mobile users are 1)syntax conversion, 2)bitstream scaling and 3)multi-modal conversion. For bitstream scaling, reduction in spatial resolution and bit-color depth are the most significant type.

Contents:
- Universal Multimedia Access
  - Concept and prior work
  - Application environments
  - Application services
- Media Conversions
  - Conventional Video Bit-rate Reduction
  - Conversions for Mobile Environment

[Wang03]
Ming-Yu Wang, Xing Xie, Wei-Ying Ma, HongJiang Zhang: "**MobiPicture: browsing pictures on mobile devices**". ACM Multimedia 2003: pp. 106-107, 2003.

This paper proposes four applications for browsing pictures on mobile devices based on image attention model (Liu03). For thumbnail views of images, instead of directly down-scaling them, smart thumbnail view crops less informative regions and keep the most informative part of the image before shrinking and down-scaled to be shown as the thumbnail which greatly improves the users' perception. For zooming some focused region, Automatic zooming-in functionality zooms the most informative part of the image. For browsing the large image in a small display, instead of scrolling and zooming manually, the optimal browsing path is generated based on the image attention model. The last application is an automatic wall-paper creation for mobile users from a large image.

[Wobbrock02]
J. Wobbrock, J.Forlizzi, S. Hudson, B. Myres: "**WebThumb: Interaction Techniques for Small-Screen Browsers**", UIST'02, pp.205-208, 2002.

This paper presents techniques for browsing web contents in small screens. A thumbnail of the screenshot of the original web content is presented to the browser to enable the user understand the whole structure of the page and select the desired part. Any part of the thumbnail can be zoomed to get more detailed information. Automatic selection and presentation of the potential interesting parts and RSVP techniques are also introduced.

[Lei03]
Z. Lei, N.D. Georganas, **"A rate adaptation transcoding scheme for real-time video transmission over wireless channels"**, Signal Process.: Image Commun. 18(8) pp. 641–658, 2003.

[Lum03]
W.Y. Lum, F.C.M. Lau, **"A context-aware decision engine for content adaptationn"**, IEEE Pervasive Comput. 1(3) 41–49, 2002.

[Good99]
N. Good, J.B. Schafer, J.A. Konstan, A. Borchers, B. Sarwar, J. Herlocker, and J. Riedl, "**Combining collaborative filtering with personal agents for better recommendations**," in Proc. Conf. Am. Assoc. Artificial Intelligence (AAAI-99), Orlando, FL, July 1999, pp. 439-446.

[Angelides03]
Marios C. Angelides, "**Multimedia Content Modeling and Personalization**", IEEE Multimedia 2003. Vol. 10, No.4, pp.12-15, 2003.

[Opera]
Opera's small screen rendering,http://www.opera.com/products/smartphone/smallscreen/

[Kim03]
J.-G. Kim, Y. Wang, S.F. Chang, "**Content-Adaptive Utility-Based Video Adaptation**", IEEE ICME-2003. July 6-9, 2003.

[Frayling02]
Frayling et.al,.: "**SmartView: Enhanced Document Viewer for Mobile Devices**", Microsoft Corp. MSR-TR-2002-114.

[Chen03]
Yu Chen, Wei-Ying Ma, Hong-Jiang Zhang: "**Detecting web page structure for adaptive viewing on small form factor devices**", World Wide Web Conference 2003, pp. 225-233, 2003.

[Kamba96]
T. Kamba, S.A. Elson, T. Harpold, T. Stamper and P.N. Sukaviriya, "**Using small screen space**

more efficiently”, CHI96 in ACM, Vancouver, Canada, 1996, pp.383-390.

[Xie04]
X. Xie, W.-Y. Ma, H.-J. Zhang, Microsoft Research Asia, China: “**Maximizing Information Throughput for Multimedia Browsing on Small Displays“**, ICME 2004 special session on Content Understanding and Transcoding Techniques for Media Adaptation.

[Divakaran04]
A. Divakaran, A. Vetro, and Takashi Kan,Mitsubishi Electric Research labs: **“Towards Maximizing the End-User Experience”,** ICME 2004 special session on Content Understanding and Transcoding Techniques for Media Adaptation.

[Thang04]
T.Thang, Y.Jung and Y. Ro: “**Modality conversion for universal multimedia services**”, WIAMIS 2004.

[Lim04]
J. Lim, S. Kang and M.Kim: “**User Preference based information personalization for easy access to multimedia contents**”, WIAMIS 2004.

[Neve04]
W. Neve, F. Keukelaere, K. Wolf, R. Walle: “**Applying MPEG-21 BSDL to the JVT H.264/AVC specification in MPEG-21 session mobility scenarios**”, WIAMIS 2004.

[Dogan04]
S.Dogan, S.Eminsoy, A. Sadka, A. Kondoz: “**Video Content Adaptation using transcoding for enabling UMA over UMTS**”, WIAMIS 2004.

[Boll99]
S. Boll, W. Klas, J. Wande: “**A Cross-Media Adaptation Strategy for Multimedia Presentations**”, ACM Multimedia 99, pp. 37-46, 1999.

References on description

[PBeek03]
van Beek, P.; Smith, J.R.; Ebrahimi, T.; Suzuki, T.; Askelof, J.; "**Metadata-driven multimedia access",** Signal Processing Magazine, IEEE , Volume: 20 Issue: 2, Page(s): 40 -52, March 2003

Abstract:

This paper presents UMA tools and applications which are supported by the use of metadata; personalization, variation, summarization and transcoding hints.

For (personalized) content selection, one or more of 1) multimedia content descriptions, 2) user preference descriptions, and 3) content usage history is used. Multimedia content descriptions allows the user to query the system and to search for desired content based on attributes like author, title, genre, language, keywords, etc. Information filtering, which utilizes a user profile to capture long-term preferences, and collaborative filtering, which applies to communities of users that share their explicit opinions or ratings of content items, is introduced to enable automatic filtering or recommendation services. MPEG-7 UsageHistory DS and UserPreference DS support these filtering functionalities.

To deliver the selected content to the user, the variation which adapts to the client terminal capabilities or user preferences should be selected or created (transcoded, summarized, etc.). MPEG-7 Variation Tools enables to describe a single content using various spatial and temporal resolution, quality, coding format, bit rate, color detail, length and modalities (video/image/audio/text). MPEG-7 media transcoding hints allow content servers, proxies, or gateways to adapt AV contents to different network conditions, user and publisher preferences, and capabilities of terminal devices with limited resources. Transcoding hints can be used for complexity reduction as well as for quality improvement in the transcoding process. MPEG-7 summary descriptions defines the summary content, how it relates to the original content, and how an actual summary of the original content can be composed from these and presented to the user.

Contents:

- Personalized selection of Multimedia Contents
    Personalization approaches, Personalized tools in MPEG-7, Application Scenarios
- Multimedia content variations
    MPEG-7 Variation Tools, Example of a variation set
- Summarization of Audiovisual Content
    Automatic summarization techniques, shot boundary detection and keyframe extraction, MPEG-7 summary descriptions, application scenarios
- Transcoding of Audiovisual Content
    Transcoding Hints in MPEG-7, transcoding of JPEG2000 imagesm, Image transcoding optimization, application scenarios

[Vetro04]
A.Vetro, "**MPEG-21 digital item adaptation: enabling universal multimedia access**", pp.84-87

This paper summarizes the specification MPEG-21 Part-7 Digital Item Adaptation.

[Li98]
C.-S. Li, R. Mohan and J. R. Smith, "**Multimedia Content Description in the InfoPyramid**," IEEE Proc. Int. Conf. Acoust., Speech, Signal Processing (ICASSP) , Seattle, WA, Special session on Signal Processing in Modern Multimedia Standards, May, 1998.

[Nastaran03]
Nastaran Fatemi, "**A semantic views model for audiovisual indexing and retrieval",** PhD Thesis N-2738, EPF Lausanne, March 2003.

[Fatemi03]
N. Fatemi, O. Abou Khaled, and G. Coray, "**An XQuery Adaptation for MPEG-7 Documents Retrieval**", XML Conference & Exposition 2003, December 2003.

This paper modifies the W3C XQuery to adapt search through TV news program archives and proposes as "Semantic Views Query Language" (SVQL). SVQL is designed to retrieve audiovisual documents described by MPEG-7 XML instances. The query language are based on five "Semantic Views" from the viewpoint of various users (creator, consumer, editor, etc). The semantic views are ***"Physical View", "Production View", "Thematic View", "Visual View", "Audio View".*** *e.g. Find: A news item in the context of Euro 2000 football games containing a shot of at least 5 seconds showing a French football supporter saying << que le meilleur gagne >>".* The implementation as a TV news querying interface is also presented.

[Fatemi01]
N. Fatemi and O. Abou Khaled, "**Indexing and Retrieval of TV News Programs Based on MPEG-7**", in Proceedings of the IEEE International Conference on Consumer Electronics (ICCE'2001), June 2001

This paper proposed a model to describe TV news contents based on MPEG-7 based on analysis of real broadcast situations. It also presents an indexing and retrieval framework.

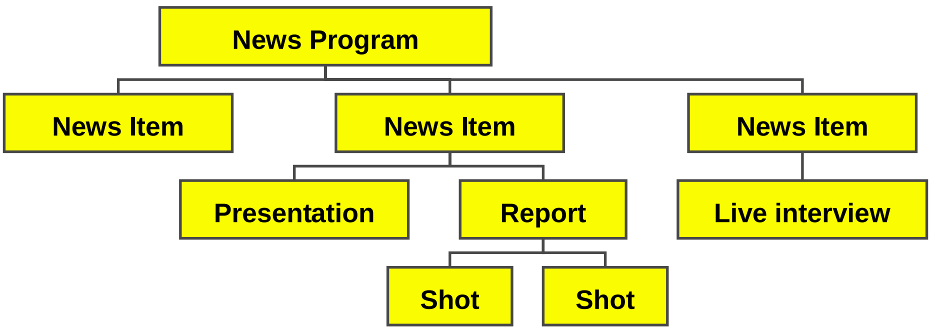


Figure: TV news content description model

Table: MPEG-7 Description Schemes for TV news content description.

TV news model	MPEG-7 DS
Video segments (news program, news item, presentation, report, etc.)	VideoSegment DS
News hierarchical structure	SegmentDecomposition DS
Script	RelatedMaterial DS
Teletext subtitles, graphical overlay	VideoText DS
Free text description	TextAnnotation DS
Keywords descriptions	StructuredAnnotation DS

[MPEG7MDS]
ISO/IEC 15938-5: “**Multimedia Content Description Interface Part 5: MDS**”, (2002).

[MPEG7Visual]
ISO/IEC 15938-3: **“Multimedia Content Description Interface Part 3: Visual**”, (2002).

[MPEG7]
B.S. Manjunath, P. Salembier, T. Sikora, **“Introduction to MPEG-7: Multimedia Content Description Language”**, Wiley, 2002.

[MPEG21]
ISO/IEC 21000-1 FCD: “**Multimedia Framework - Part 1: Vision, Technologies and Strategy**”, (2003).

[MPEG21DID]
ISO/IEC 21000-2 DIS: “**Multimedia Framework - Part 2: Digital Item Declaration**”, (2003).

[MPEG21IPMP]
ISO/IEC 21000-4 CD: “**Multimedia Framework - Part 4: Intellectual Property Management and Protection**”, (2003).

[MPEG21DIA]
ISO/IEC 21000-7 DIS: “**Multimedia Framework - Part 7: Digital Item Adaptation**”, (2003).

[CC/PP]
Composite Capability/Preference Profiles (CC/PP): Structure and Vocabularies 1.0, 2004.

[UAProf]
User Agent Profiling Specification, http://www.openmobilealliance.com/

[PSS232]
3GPP TS 22.233: "**Transparent End-to-End Packet-switched Streaming Service; Service aspects; Stage 1"**.

[PSS233]
3GPP TS 26.233: "**Transparent end-to-end packet switched streaming service (PSS); General description**".

[PSS234]
3GPP TS 26.234: "**Transparent end-to-end packet switched streaming service (PSS); Protocols and codecs V5.6.0**", 2003.

[DI]
"**Authoring Challenges for Device Independence**", W3C Working Group Note 1 September 2003 ,http://www.w3.org/TR/acdi/

[TVA]
TV Anytime Forum. http://www.tv-anytime.org/

[DCORE]
DublinCore, http://dublincore.org/

[SMTPE]
SMPTE Metadata Dictionary as specified in SMPTE RP210.4, http://www.smpte-ra.org/mdd/index.html, 2003

[PMeta]
P/Meta, http://www.ebu.ch/

[MXF]
Metadata eXchange Format (MXF), http://www.pro-mpeg.org/

[AAF]
AAF specification v.1.0.1 (draft) , http://aafassociation.org/, 2003.

[SCORM1.2]
"Sharable Content Object Reference Model (SCORM) Version 1.2", ADL Technical Team, October 2001.

[LOM]
Standard for Information Technology --Education and Training Systems -- Learning Objects and Metadata, http://ltsc.ieee.org/wg12/

[NewsML]
NewsML 1.2,
http://www.newsml.org/IPTC/NewsML/1.2/documentation/NewsML_1.2-doc-Guidelines_1.00.pdf
2004.

References on system designing and other

[Bormans03]

Bormans, J.; Gelissen, J.; Perkis, A.; “**MPEG-21: The 21st century multimedia framework”,** Signal Processing Magazine, IEEE , Volume: 20 Issue: 2 , Page(s): 53 -62, March 2003.

This paper describes how MPEG-21 digital item adaptation (DIA) can give solutions for UMA after outlining the context and background of the MPEG-21 initiative and an overview of MPEG-21 technology. UMA deals with the delivery of the media resources under different network conditions, User preferences, and capabilities of terminal devices. UMA presents the solution for wired and wireless systems to access the same media resource provider, each of them receiving media resources. MPEG-21 Part-7 DIA is designed to adapt the DIs (Digital Items) according to the actual usage environment and the media resource adaptability, including content representation format and resource complexity descriptions. Main factors are content availability, terminal capabilities, network characteristics, user preferences, natural environment of the user, and the streaming media characteristics of these factors are available bandwidth, error characteristics, screen size, content scalability (spatial,temporal, and spectral), adaptavility, interactivity, synchronization and multiplexing.

[Chung03]

Chung-Neng Wang; Chia-Yang Tsai; Hsiao-Chiang Chuang; Yao-Chung Lin; Jin-He Chen; Kin Lam Tong; Feng-Chen Chang; Chun-Jen Tsai; Shuh-Ying Lee; Tihao Chiang; Hsueh-Ming Hang; **"FGS-based video streaming test-bed for MPEG 21 universal multimedia access with digital item adaptation",** Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on , Volume: 2 , Page(s): II-364 -II-367 vol.2, 25-28 May 2003.

This paper presents the system architecture of a video streaming prototype system used as a reference test bed of MPEG-21 DIA. This system streams real-time video over heterogeneous networks to devices with different capabilities using MPEG-4 Fine Granularity Scalability(FGS) profile and MPEG-21 DIA.

Contents:

- FGS-based streaming test bed
  - FGS-based Video Content Server
  - Video Clients
  - Network Interface
  - Network Similator
- Experimental Results

[Etoh02]

M. Etoh, S. Sekiguchi, “**MPEG-7 enabled Digest Video Streaming over 3G Mobile Network**”, Packet Video 2002.

This paper proposes a framework for personalized video digest delivery over wireless networks. Terminal-side content adaptation mechanism by using the concept of mobile agent is employed to generate personalized video digests dynamically taking user's preferences and situations regarding content viewing into account. A mobile agent controls the interaction among metadata sets that describe the mobile video delivery environment. MPEG-7 standard is used to describe video structures.

[Fossbakk01]

Fossbakk, E.; Manzanares, P.; Yago, J.L.; Perkis, A.; "**An MPEG-21 framework for streaming media",** Multimedia Signal Processing, 2001 IEEE Fourth Workshop on , Page(s): 147 -152, 3-5 Oct. 2001.

Abstract:

This paper presents the experimental model for simulating an MPEG-21 framework for streaming media. The latest version encodes video sequences in MPEG-4 with temporal, spatial and FGS scalability, DMIF based streaming, variation selection engine which selects the best variation from the viewpoint of the available bandwith of the user, and network emulator between the server and client which emulates latency, probability of congestion, link fault and rate of packet loss.

Contents:

- Digital Item Declaration
- Building the experimental test bed
    - Concept demonstrator
    - Integration of a server
    - the multimedia framework

[Pereira03]

Pereira, F.; Burnett, I.; **"Universal multimedia experiences for tomorrow",** Signal Processing Magazine, IEEE , Volume: 20 Issue: 2 , Page(s): 63 -73, March 2003.

This paper discusses the current status of universal multimedia access (UMA) technologies and investigates future directions in this area. Key developments and trends from the last few years have set the scene for ubiquitous multimedia consumption. In summary, these are: wireless communications and mobility; standardized multimedia content; interactive versus passive consumption; and the Internet and the World Wide Web (WWW). However, the most relevant emerging trend is that the end point of universal multimedia consumption is the user and not the terminal. Therefore, the vision of mass delivery of identical content is being replaced by one of mass customization of content centered on the user.

Contents:

- Recent Key Developments
    - Wireless communications and mobility

- Standardized multimedia content
    - Interactive versus passive consumption
    - Internet and WWW
- The UMA problem (lack of standard technologies, network and terminal bottlenecks)
- From Access to Experience
    - Experiences and Knowledge
    - Senses and sensors
    - Experience Limitations
- Emerging and future trends and technologies
    - Existing and emerging technologies
    - Content Representation
    - Content Representation (Scalable Coding, Transcoding)
    - User Environment Description
    - Content Adaptation
    - Intellectual Property Management and Protection (content and user profile)
    - Presentation Conditions and Devices
    - Multiple Terminals at Work
    - Mobile and Wearable Devices
    - Active and Programmable Networks
    - Peer-to-peer Content Delivery
    - Role of open standards
- Limitations and Risks
- Final Remarks

[Steiger03]
Olivier Steiger, David Marimón Sanjuán and Touradj Ebrahimi, "**MPEG-Based Personalized Content Delivery,**", Proc. of the IEEE Int. Conf. on Image Processing, ICIP 2003,
Barcelona, Spain, September 14-17, 2003.

A personalized multimedia content delivery system using user preferences and terminal/network capabilities are presented. Key issues of the system are content preparation (variation, MPEG-7 annotation tool), content adaptation and delivery using user/server preferences, terminal/network capabilities and usage history using MPEG-7 and MPEG-21 descriptions.

[Vetro03]
Vetro, A.; Christopoulos, C.; Ebrahimi, T.; "**From the guest editors - Universal multimedia access",** Signal Processing Magazine, IEEE , Volume: 20 Issue: 2 , Page(s): 16 -16, March 2003.

Abstract:
This paper gives the abstract of the necessity of UMA and its key issues. UMA refers to the framework where information is accessed in a suitable form and modality under the current complex and dynamic usage environment such as devices, networks, terminals, user preference,

personalization, and other factors of usage environment. To adapt the content accordingly, content representation format, description and management of content is important. The development of a standardized description of usage environments to enable, for example, negotiation of device characteristics and QoS parameters is also important, and Digital item adaptation is a key component in MPEG-21 to this aspect. All of these aspects are included in IEEE Signal Processing Magazine 2003.3.

[Perkis01]
A. Perkis, Y. Abdeljaoued, C. Christopoulos, T. Ebrahimi, and J.F. Chicharo, "**Universal multimedia access from wired and wireless systems,**" Birkhauser Boston trans. on Circuits, Systems and Signal Processing, vol. 10, no. 3, pp. 387–402, 2001.

[Masumitsu01]
Ken Masumitsu, Tomio Echigo: "**Metadata Framework for Constructing Individualized Video Digest** ", ICIP 2001, pp.390-393, 2001

This paper presents a framework for providing video digests that are personalized by profiles of individual users. Video contents have meta-data described manually from a set of predefined keywords that have temporal duration. Content profiles are prepared by a provider, which are vectors of the importance value of keywords, and only one should be selected by a user. In addition, a user profile is collected by the user, which has the same components. The importance scores of an image sequence along the time axis can be calculated from a combination of these profiles. Finally, the video clips can be collected as the video digest from the whole contents, which have higher importance scores than a threshold transformed from the length of the user requirement.

[Sampath00]
L.Sampath, A.Heral, J.R.Smith, "**UMA-based Wireless and Mobile Video Delivery Architecture**", Proc. SPIE Vol. 4210 Internet Multimedia Management Systems, pp. 103-115, 2000.

This paper presents a framework for describing the network conditions under which the multimedia data is being transmitted, as well as the modality of the device which receives the data. This presents an architecture for adaptation of multimedia data to such wireless network conditions and device capabilities, under constraints imposed by user preferences and multimedia content, to ensure effective, meaningful, and acceptable delivery of video data to mobile users. The adaptability is achieved through careful application of a combination of off-line and on-line reductions to the video streams.

The architecture consists of an MPEG-7 server in the fixed network, and an MPEG-7 player on the mobile host. In addition, an encoder/decoder layer is used to perform physical frame bit reductions on transcoded video frames. The server also maintains index structures and a programmable two-dimensional matrix of reductions. An offline packager module is included with the server, which packages all the MPEG-7 data with the video when the video is first registered with the database.

[Bickmore97]
T. W. Bickmore and B. N. Schilit, “**Digestor: Device-Independent Access to the World Wide Web**”, Sixth World Wide Web Conference, April 7-11, 1997, pp. 655-663, 1997.

[Magalhaes04]
J. Magalhães, F. Pereira, "**Using MPEG standards for multimedia customization**", Signal Processing: Image Communication, 2004.

[Nagao01]
K. Nagao, “**Semantic Annotation and Transcoding: Making Web Content More Accessible**”, IEEE MultiMedia, Vol.8, No.2, pp.69-81, 2001

[Nagao03]
K. Nagao, “**Digital Content Annotation and Transcoding**”, Artech House Publishers, 2003

[Tseng04]
B.L.Tseng, Ching-Yung Lin, J.R. Smith, “**Using MPEG-7 and MPEG-21 for personalizing video**”, pp. 42- 52

[Herranz04]
L. Herranz and Jose Martinez: “**Towards universal and personalized access to audiovisual content in the DYMAS system**”, WIAMIS 2004.

[Lankhorst02]
M. Lankhorst, H. Van Kranenburg, A. Salden, A. Peddemors, "**Enabling Technology for Personalizing Mobile Services**", IEEE proc. of HICSS'02-Volume 3, 2002.

[Shen04]
B. Shen, Z. Xu, S. Wee, J. Apostolopoulos, HP Labs, California: “**Semantic-Enhanced Distribution and Adaptation Networks”**, ICME 2004 special session on Content Understanding and Transcoding Techniques for Media Adaptation.

[Lin04]
C.-Y. Lin, A. Natsev, B. L. Tseng, M. Hill, J. R. Smith, C.-S. Li, IBM T. J. Watson Center: “**Content Transcoding Middleware for Pervasive Geospatial Intelligence Access**”, ICME 2004 special session on Content Understanding and Transcoding Techniques for Media Adaptation.

[Laszlo03]
László Böszörményi, Hermann Hellwagner, Harald Kosch, Mulugeta Libsie, Stefan Podlipnig, “**Metadata Driven Adaptation in the ADMITS Project”**, EURASIP Signal Processing: Image Communication Journal, Vol. 18, No. 8, Sept. 2003 (Special Issue on Multimedia Adaptation), pp. 749-766

[Panis03]
Gabriel Panis, Andreas Hutter, Jörg Heuer, Hermann Hellwagner, Harald Kosch, Christian Timmerer, Sylvain Devillers and Myriam Amielh: "**Bitstream Syntax Description: A Tool for Multimedia Resource Adaptation within MPEG-21",** EURASIP Signal Processing: Image Communication Journal, Vol. 18, No. 8, Sept. 2003 (Special Issue on Multimedia Adaptation), pp. 721-747

[Timmerer03]
C. Timmerer, G. Panis, H. Kosch, J. Heuer, H. Hellwagner, A. Hutter
Coding format independent multimedia content adaptation using XML
In: Proceedings of SPIE International Symposium ITCom 2003 on Internet Multimedia Management Systems IV, Vol. 5242
Orange County Convention Center, Orlando, Florida, USA, 7-11 September 2003

[Kosh03]
Harald Kosch: "**Distributed Multimedia Database Technologies supported by MPEG-7 und MPEG-21",** CRC Press. 248 pages. November 2003. ISBN: 0-849-31854-8.

[IEEEIC01]
Special Issue on Personalization and Privacy, IEEE Internet Comput., vol. 5,pp. 29-62, Nov./Dec. 2001.

[Ma00]
W.Y. Ma, I. Bedner, G. Chang, A. Kuchinsky, H.J. Zhang, "**A framework for adaptive content delivery in heterogeneous network environments**", Proc. of the SPIE/ACM Conference on Multimedia Computing and Networking 2000, pp. 86–100, 2000,

[Kasutani04]
Eiji Kasutani, "**Investigation Report on Universal Multimedia Access**", EPFL-ITS Technical Report 04-04, January 2004.

References on privacy

- Trust:

[Nodder03]
C. Nodder (Microsoft), "**Say versus Do; building a trust framework through users' actions, not their words.**", UbiComp 2003

[Langheinrich03]
M. Langheinrich (ETHZ), "**When Trust Does Not Compute – The Role of Trust in Ubiquitous Computing**", UbiComp 2003

- Location:

[Hong03]
J.I.Hong (UCB), "**Privacy and Security in the Location-enhanced World Wide Web**", UbiComp 2003

[Myles03]
G.Myles (Arizona Univ.), "**Preserving Privacy in Environments with Location-based Applications**", IEEE Pervasive computing, 2003.

- System design, general framework:

[Schwartz00]
Ari Schwartz and Deirdre Mulligan, "Your Place or Mine?: Privacy Concerns and Solutions," Computers, Freedom and Privacy Conference (CFP), April 2000.

[Karjoth02]
G.Karjoth and M.Schunter (IBM), "**A Privacy Policy Model for Enterprises**", 15th IEEE Computer Security Foundations Workshop (CSFW'02), 2002.

[Langheinrich02]
M. Langheinrich (ETHZ), "**Privacy Invasions in Ubiquitous Computing Environments**", UbiComp 2002.

[Boyd02]
D. boyd, et. al, "**Privacy in digital environments: Empowering users.**", Workshop abstract in Extended Abstracts of the ACM Conference on Computer Supported Cooperative Work (2002).

[Langheinrich-Ubicomp02]
M. Langheinrich (ETHZ), "**A Privacy Awareness System for Ubiquitous Computing Environments**", UbiComp 2002.

[Ackerman99]
M. Ackerman (UCB), "**Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences**", ACM Conf. E-Commerce, 1999.

[Bellotti93]
R. Bellotti, "**Design for Privacy in Ubiquitous Computing Environments**", ECSCW93, 1993.

[CDT]
CDT's guide to online privacy, http://www.cdt.org/privacy/guide/introduction/

[Koizumi02]
Y. Koizumi, "**Personal information protection in broadband environment**." (in Japanese), 2002 NEC Planning Research, Ltd., 2002.
http://www.nepr.co.jp/groups/research/social/pdf/broadband-2002-02.pdf

- User behavior analysis:

[Beckwith03]
R. Beckwith (Intel), "**Designing for Ubiquity: The Perception of Privacy**", IEEE Pervasive computing, 2003

1. Capability Configuration, harmonization between CC/PP and P3P:

[Oshbakk02]
P. Osbakk and N.Ryan. "**Context, CC/PP, and P3P**", UbiComp 2002 Adjunct Proceedings, pages 9-10, Göteborg, Sweden, September 2002.

[Okada01]
W. Okada, M. Kato.(Keio univ), "**Adapt CC/PP to Real World Information to Customize Web Service**", Poster Proc., pp108-109, WWW10, Hong Kong, May 2001,

[Kato01]
M. Kato, W. Okada. (Keio univ), "**Architecture and Implementation of CC/PP Harmonization with P3P based on Web Privacy.**", Poster Proc., pp24-25 WWW10, Hong Kong, May 2001,

- Definition of terms:

[Clarke97]
R. Clarke, "**Introduction to Dataveillance and Information Privacy, and Definitions of Terms**", http://www.anu.edu.au/people/Roger.Clarke/DV/Intro.html, 1997.

[Clarke99]
Roger Clarke. Internet privacy concerns confirm the case for intervention. Communications of the ACM, Special Issue on Internet Privacy, 2(42):60–67, February 1999.

- Standards:

[Cranor02]
Lorrie Faith Cranor, "**Web Privacy with P3P**", O'Reilly & Associates, ISBN 0-59600-371-4, 2002.

[Truste]
"**The TRUSTe White Paper**", http://www.truste.org/

[P3P]
"**The Platform for Privacy Preferences 1.0 (P3P1.0) Specification**", W3C Recommendation 16 April 2002, http://www.w3.org/TR/P3P/

[Appel]
"**A P3P Preference Exchange Language 1.0 (APPEL1.0)**", W3C Working Draft 15 April 2002, http://www.w3.org/TR/P3P-preferences/

[MPEG21RDD]
"**MPEG-21 Part 5 - Rights Data Dictionary**", 2003.

[MPEG21RLL]
"**MPEG-21 Part 6 - Rights Expression Language**", 2003.

t

Appendix A: How to use the demonstrator

Screen size adaptation DEMO overview

This demonstration introduces a system for browsing video contents with their metadata adapting any screen size. The video contents and their metadata are transformed and presented in a way that maximizes the total content value within a restricted screen size.
The total content value is automatically optimzed by
1) determining the optimal balance between the video pane and the metadata pane,
2) by transforming the video content size to adapt the screen size, and
3) by controlling the number of elements, layers and attributes to be presented adapting the screen size.
This system enables users to browse video contents with their metadata optimized to the screen size of their devices.

How to use

1. Access the URL of the demonstrator. (e.g. http://itswww.epfl.ch/~eiji/umademo/)
2. Select one video content from the list.
3. Input the desired screen size.
4. Click “Edit browsing preferences” if there are browsing preferences.
5. Input the balance between video and metadata in terms of importance and layout with the box on the left checked. If the box is not checked, the adaptation process will try to maximize the video pane size.
6. Click SEND button.
7. The adapted content appears in a popup window of a given size. The user can interact with the given content for flexible video and metadata browsing.

Note: this demo requires the Real Player plug-in to work properly. Also, please make sure to set its target bitrate properly to get adequate video quality.