Eiji Kasutani

Investigation Report on Universal Multimedia Access

Supervisor: Professor Touradj Ebrahimi

ITS Report 04.04

Lausanne, EPFL
January 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 investigation report analyzes the state-of-the-art technologies in UMA and tries to identify the key issues of UMA. The state-of-the-art in multimedia content adaptation, an overview of the standards that supports the UMA framework, potential privacy problems in UMA systems and some new UMA applications are presented in this report. This report also provides challenges that still remain to be resolved in UMA to make clear the potential key problems in UMA and determine which ones to solve.

Table of contents

List of figures

List of tables

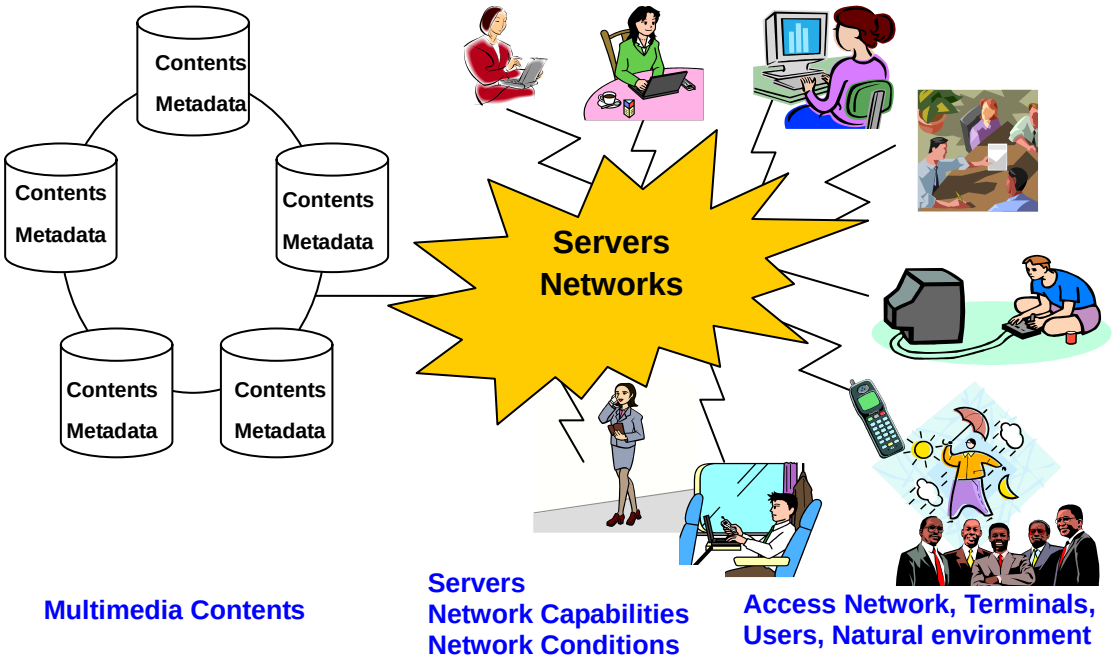

Figure 1. Multimedia content access through different environments.

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 investigation report analyzes the state-of-the-art technologies in UMA, and tries to identify the key issues of UMA and key problems that still remain to be resolved in UMA.

2. Universal Multimedia Access (UMA)

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. 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. 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, user preference, 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].

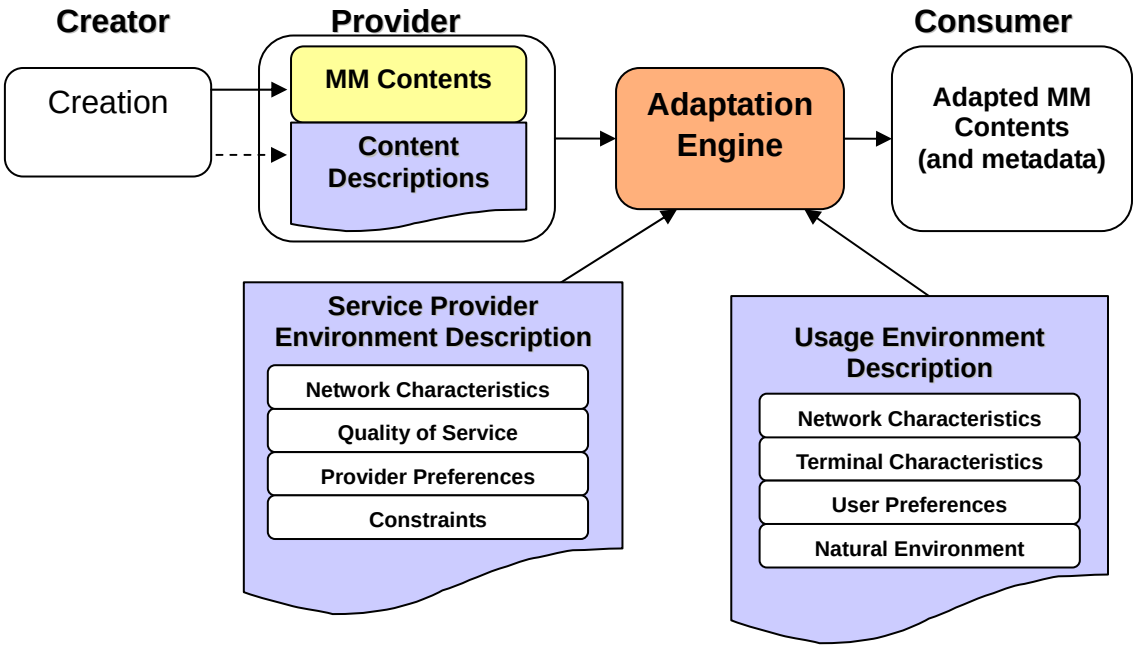

Figure 2. Overview of Universal Multimedia Access System.

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. Considering a scenario where the content is adapted to the Consumer according to these three types of descriptions, it is obvious that such a system would have a component implementing the content adaptation to the Consumer, an **Adaptation Engine**. 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, not only the adaptation engine but also the protection of both the **consumers' privacy** and the **provider and content holders' rights** of the contents are very important in UMA systems 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 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 between provider and consumer.
B) (Standardized) Description for adaptation.
C) Privacy protection of consumer (and provider).
D) Rights management of service provider and content holder.
E) UMA application and system designing.

A) Adaptation between Provider and Consumer (User-centric adaptation):

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.

B) (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.

C) 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.

D) 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][22][23].

E) UMA Application and System Design:

All of the aspects described in A) to D) should be considered for UMA systems, especially in the adaptation process. Therefore, 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. 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.

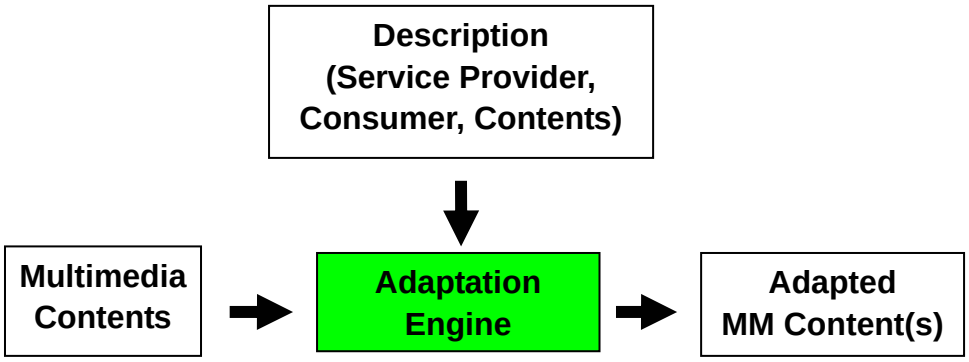

Figure 3. Content adaptation framework in UMA systems.

3. Content Adaptation in Universal Multimedia Access Systems

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.

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

*Program: 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, 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 piece of ‘program’ or various ‘programs’. We use the term ‘program’ for easy understanding considering the television program case, if there are just a single television program or there are various television programs. ‘Program’ does not refer only to television programs, but to any multimedia contents (e.g. a song). The second point is if there are variations of each program created from a single source or not. Several alternative versions or variations of a single piece of program 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 program with no variation, requires a transformation engine that transforms the target program into an adapted content. The second type, where there is a single program 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 programs with no variation, requires a program selection engine that selects the best program 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 programs with multiple variations, requires a program and variation selection engine that selects the best program and variation. After selecting the best program 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. 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:

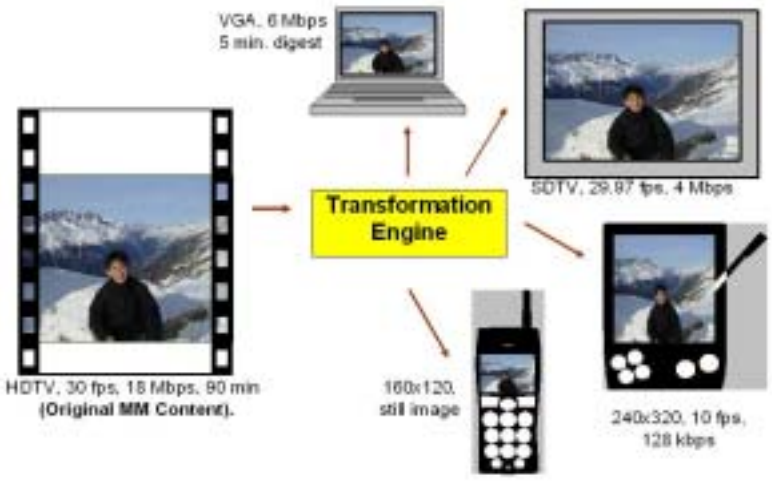

Figure 5. Example of content adaptation by transformation

- **Conversion of spatio-temporal resolution** (e.g. VGA to QVGA, 24bit -> 8bit color depth, 30fps to 10 bps).
- **Transcoding** [SumISCAS03].
  - Conversion of video coding formats (e.g. MPEG-2 to MPEG-4)
  - Coding parameter conversion.
    - ✧ 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)
- **Transmoding** (e.g. speech-to-text conversion, video-to-image conversion (mosaicing, key frame extraction), image-to-text conversion etc.) [Cavallaro03].
- **Spatial and temporal scalable coding** (e.g. MPEG-4 FGS [Chung03][ChenICME02] [Shaar02][ChenICCE02]).
- **Summarization** (e.g. 1.5 times fast play, 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.)
- (Terminal) **Capability negotiation** [12][13][14].
- **Selection of regions** (spatial (e.g. primary object selection), temporal (important video segment selection)). The use of Region of Interest (ROI) are also included.

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.

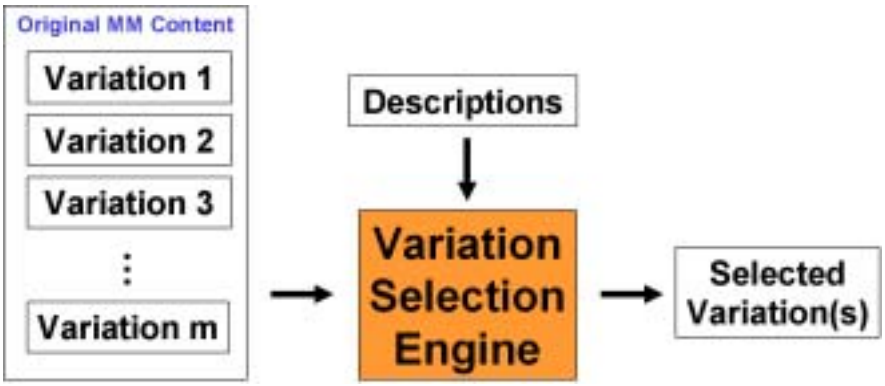

Figure 6. Content adaptation by variation selection.

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, and furthermore, 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, and color selection for the user. Some content variation selection examples are described in the following:

- 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.

- 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).

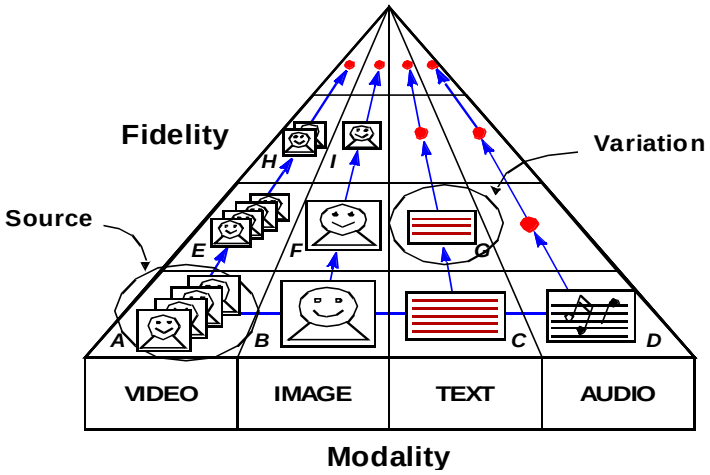

Figure 7. Illustration of different variations of a single source content.

- Selection of the best variation based on service provider preferences and content descriptions (e.g. content designed for only TV or mobile phone, limitations on access networks, terminals, etc).

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 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.

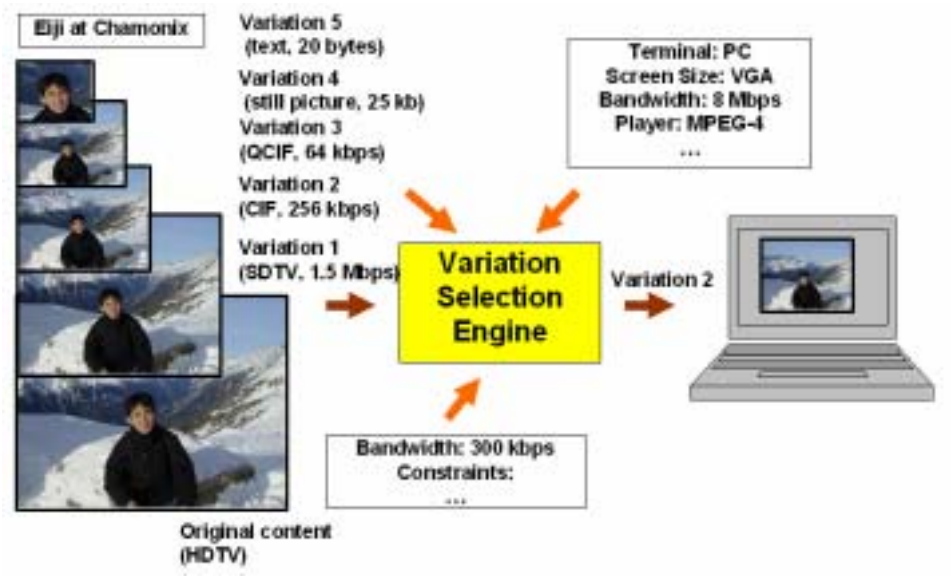

Figure 8. Example of content adaptation by variation selection.

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)

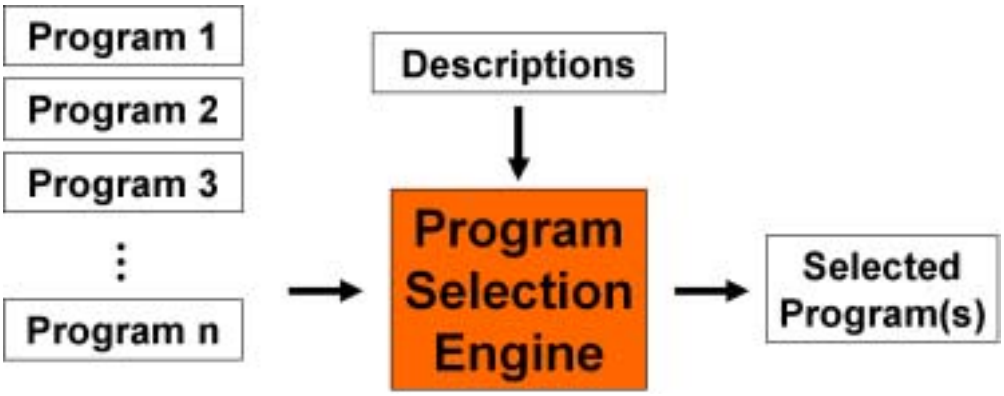

Figure 9. Content adaptation by program selection.

3.2.3. Program selection engine

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

Some program 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…).
- Automatically recording 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.

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

- Content filtering (rule-based, content-based, collaborative filtering) [Angelides03].
- 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 program(s) by the program 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 program 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?

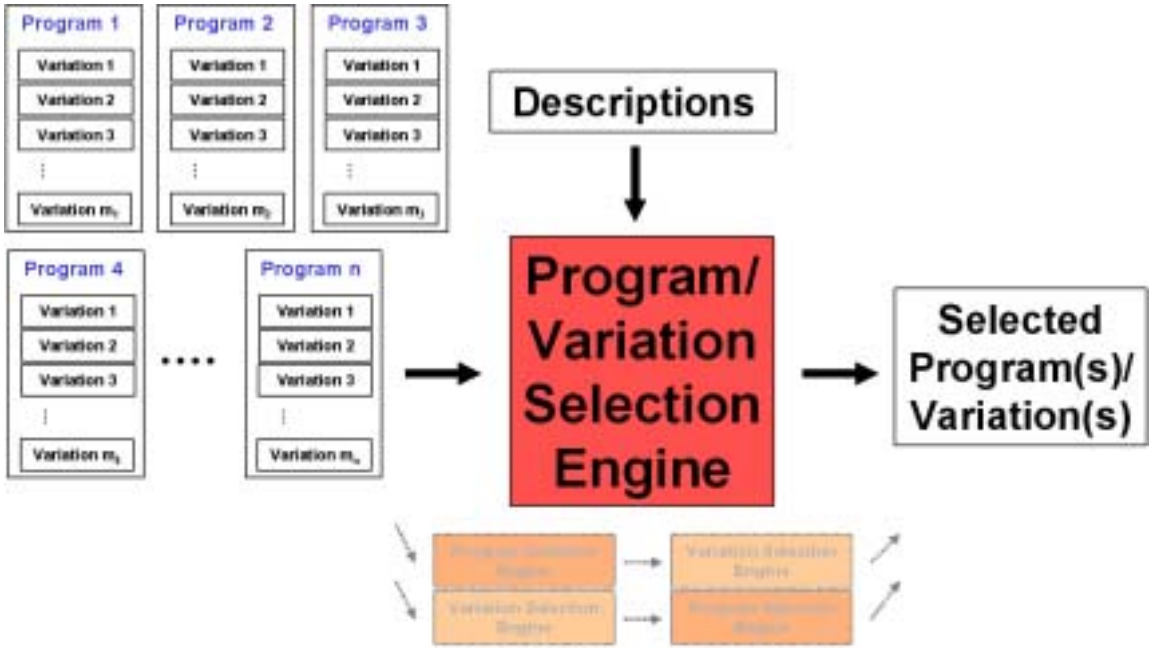

Figure 10. Content adaptation by program and variation selection.

3.2.4. Program 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.

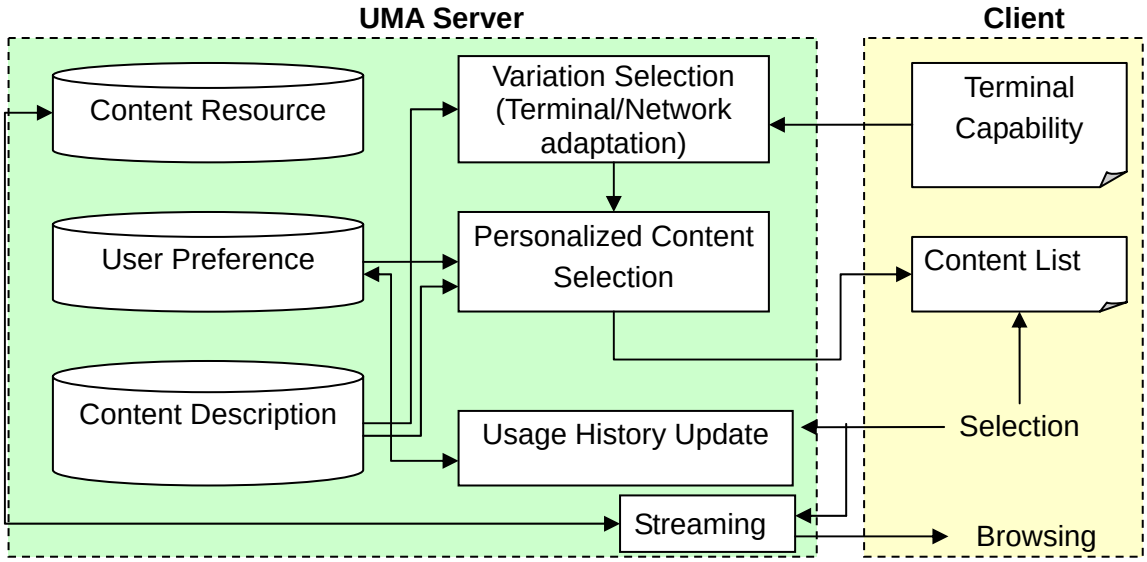

Figure 11. Personalized content delivery using program and variation selection engine.

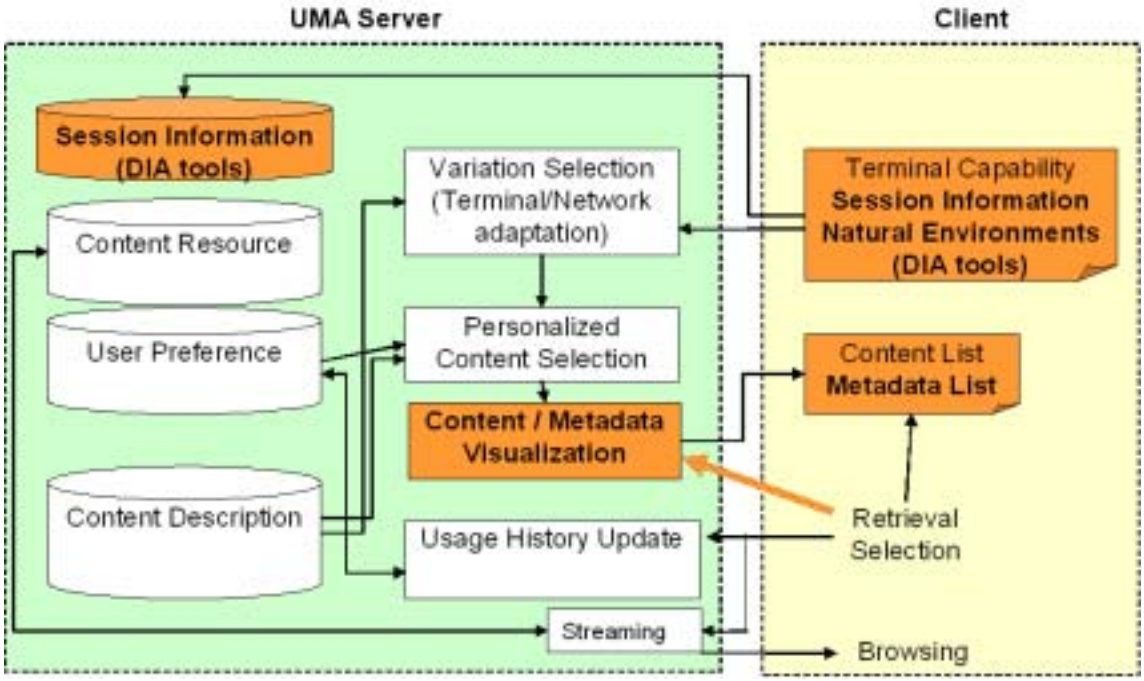

Figure 12. Example of Personalized content delivery with visualization functions.

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.

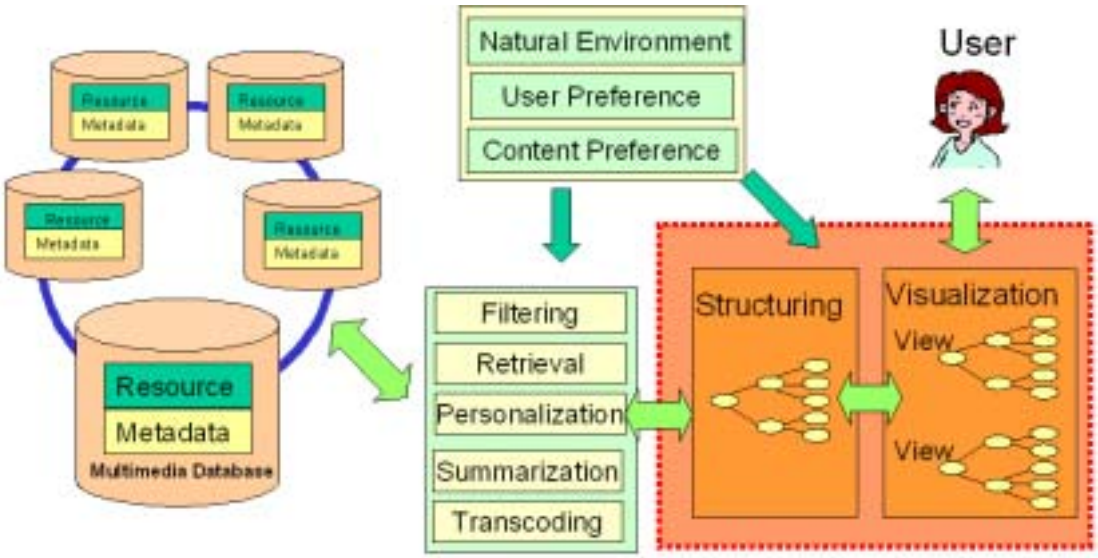

Figure 13. Metadata Visualization

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.

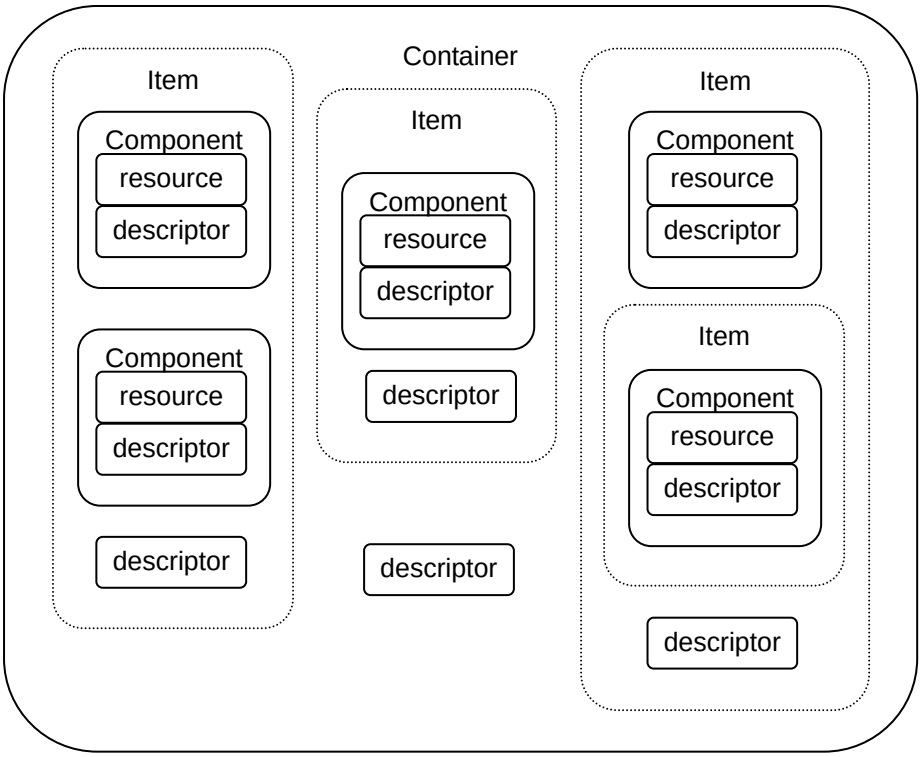

Figure 14. Relation between MPEG-21 elements

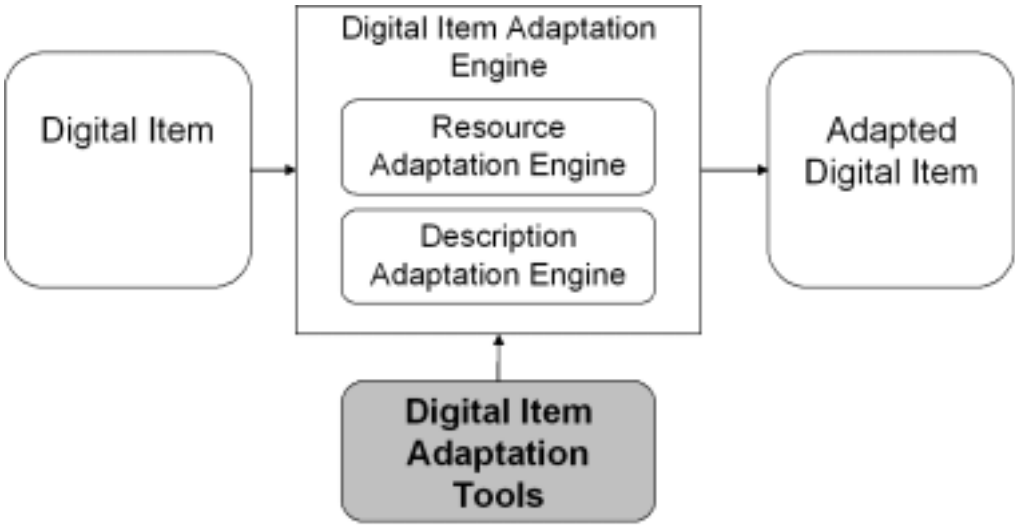

Figure 15. Illustration of MPEG-21 Digital Item Adaptation.

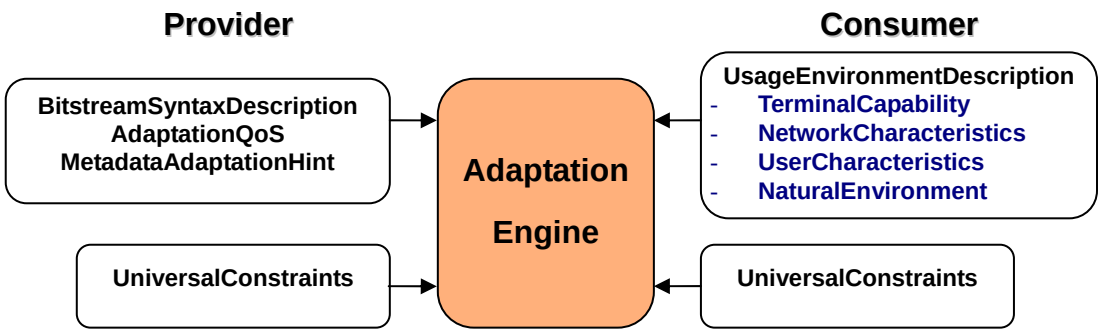

Figure 16. DIA tools in UMA systems.

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.

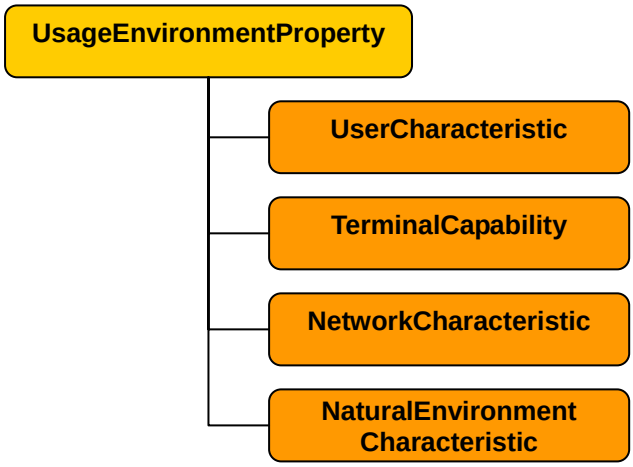

Figure 17. Usage Environment description tools

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.

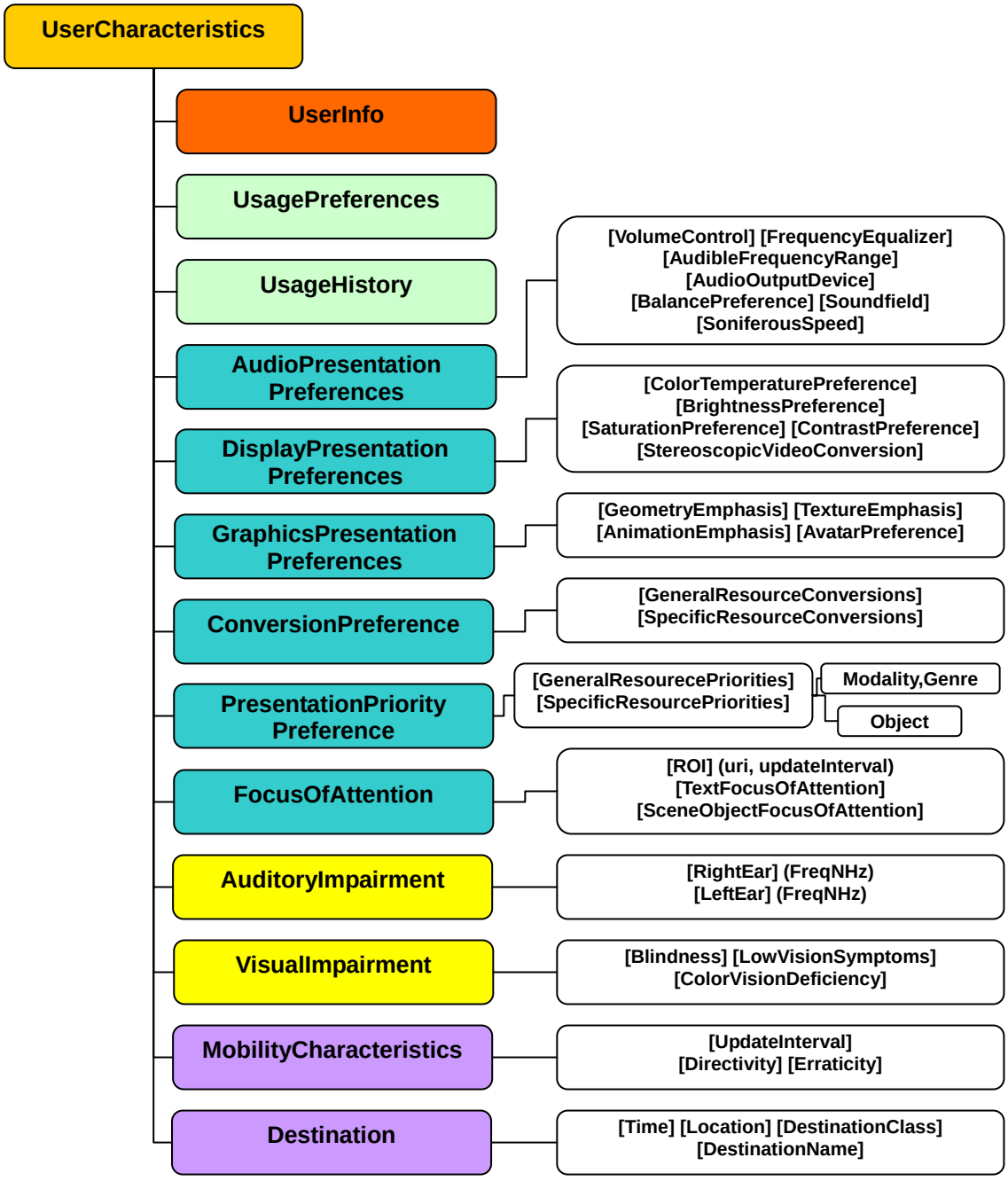

Figure 18. User Characteristic description tools

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)

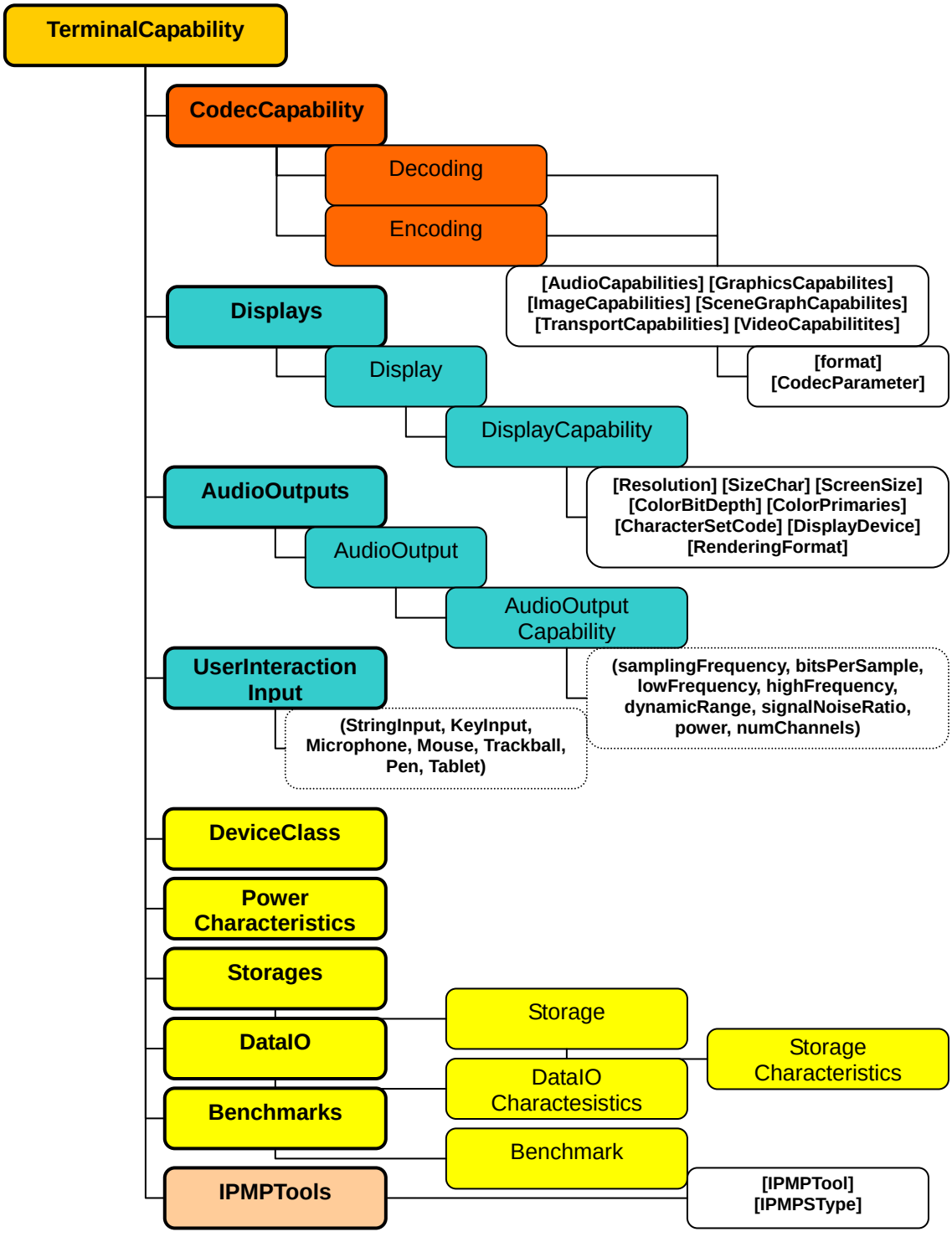

Figure 19. Terminal Capability description tools.

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..

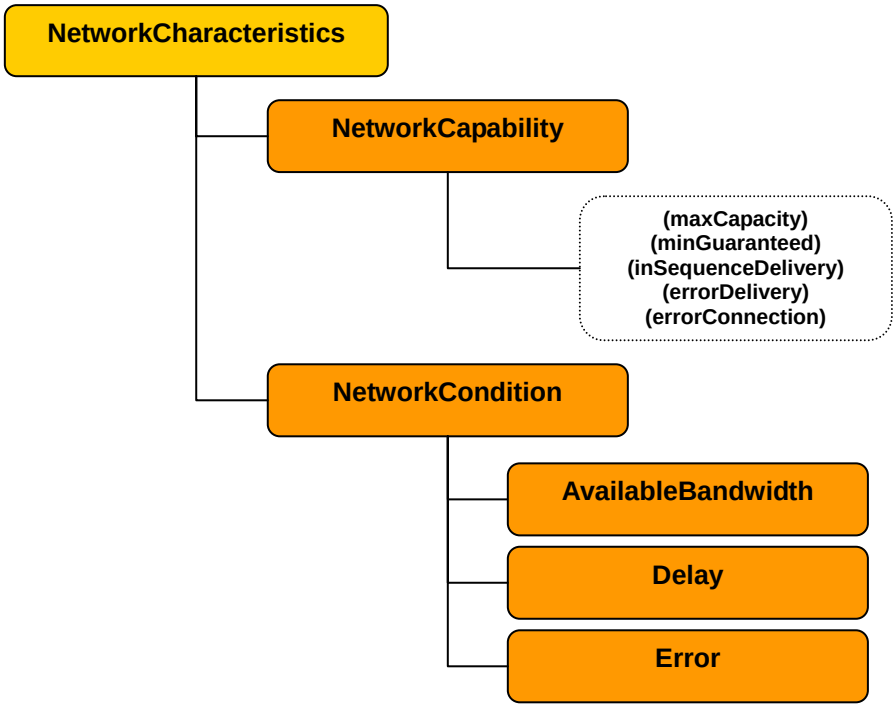

Figure 20. Network Characteristics description tools.

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.

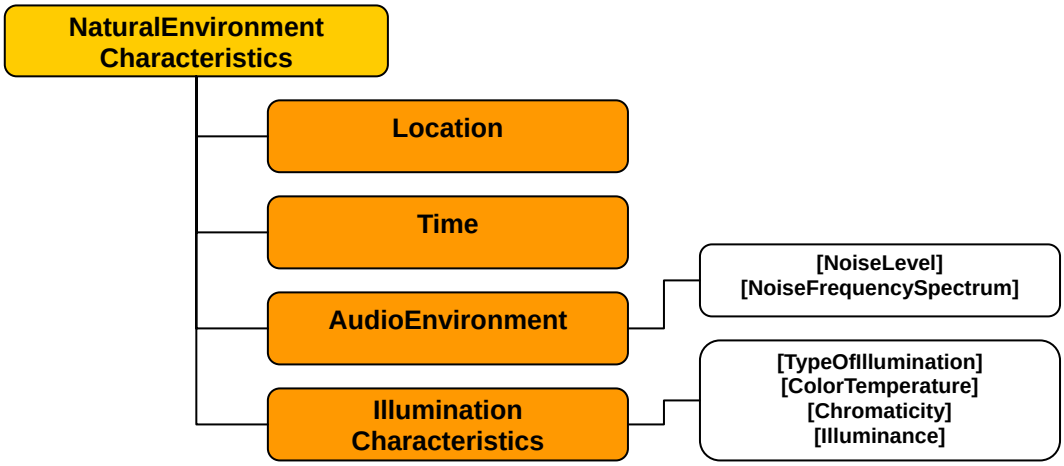

Figure 22. Natural Environment Characteristics description tools.

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.

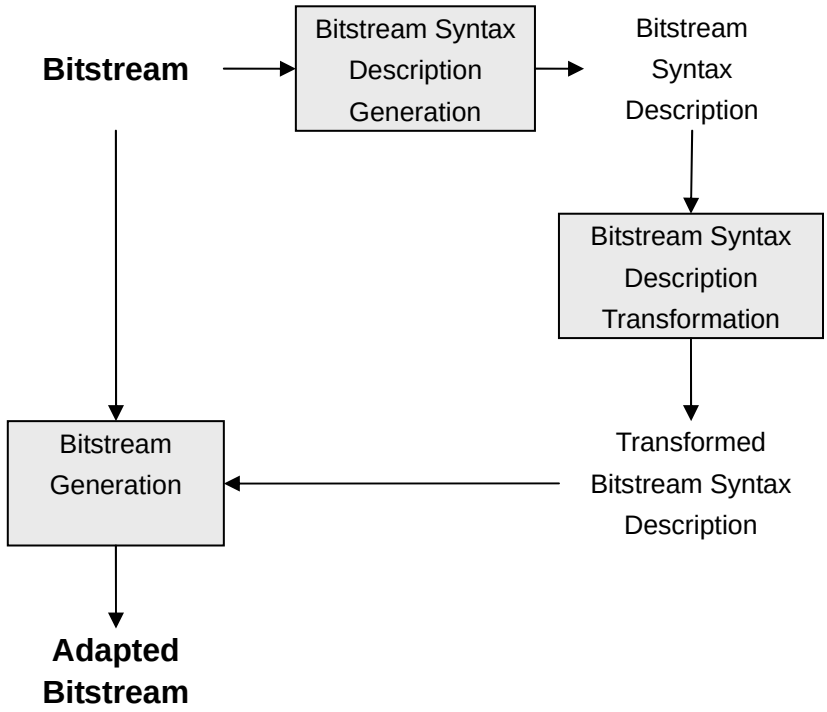

Figure 23. Adaptation architecture.

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.

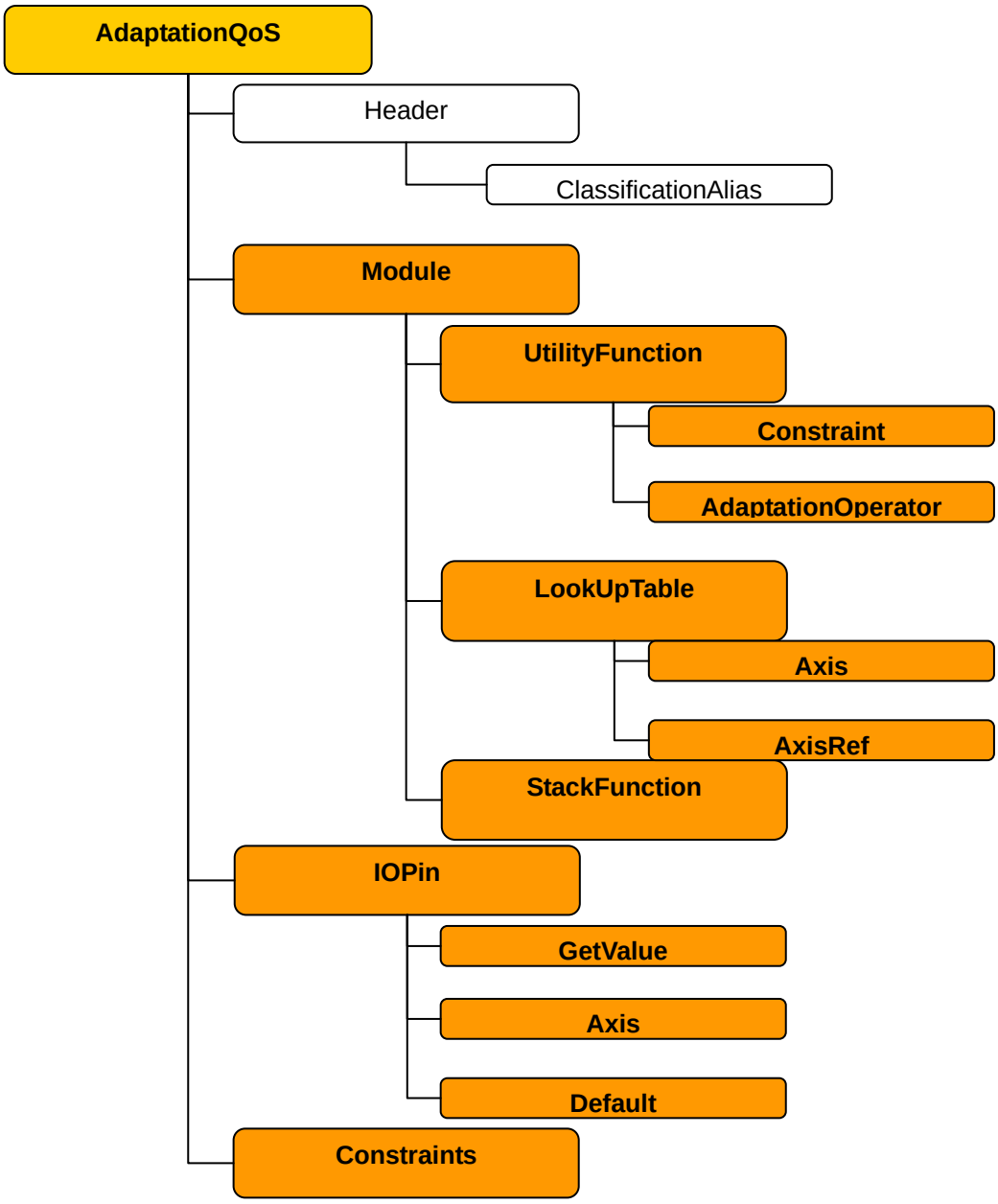

Figure 24. Terminal and Network quality of service description tools

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.

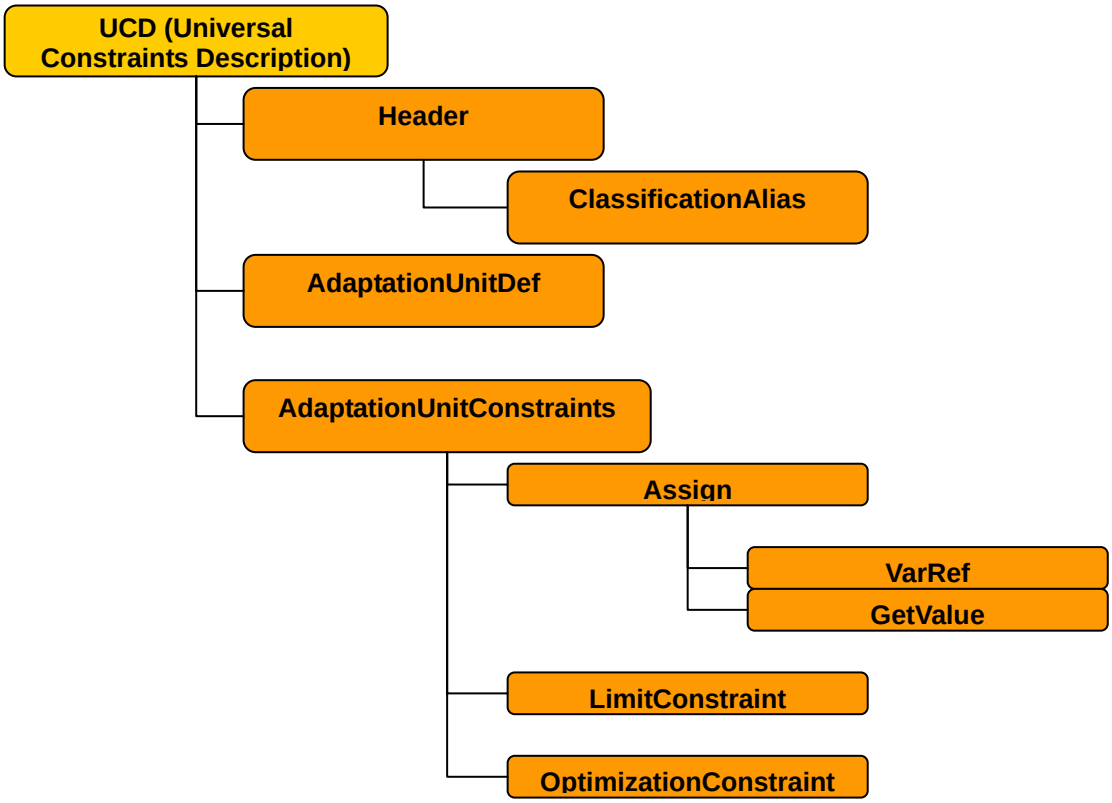

Figure 25. Universal constraints description tools

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.

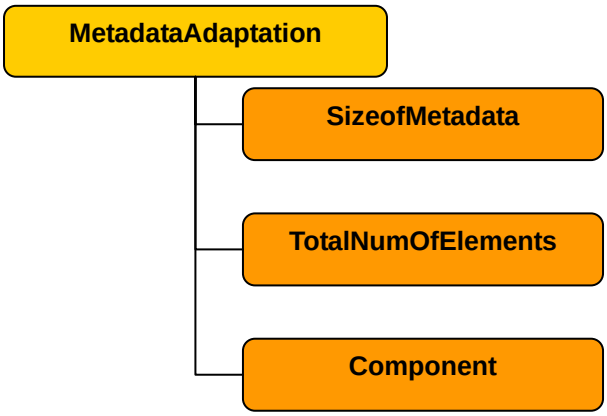

Figure 26. Metadata Adaptavility

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.

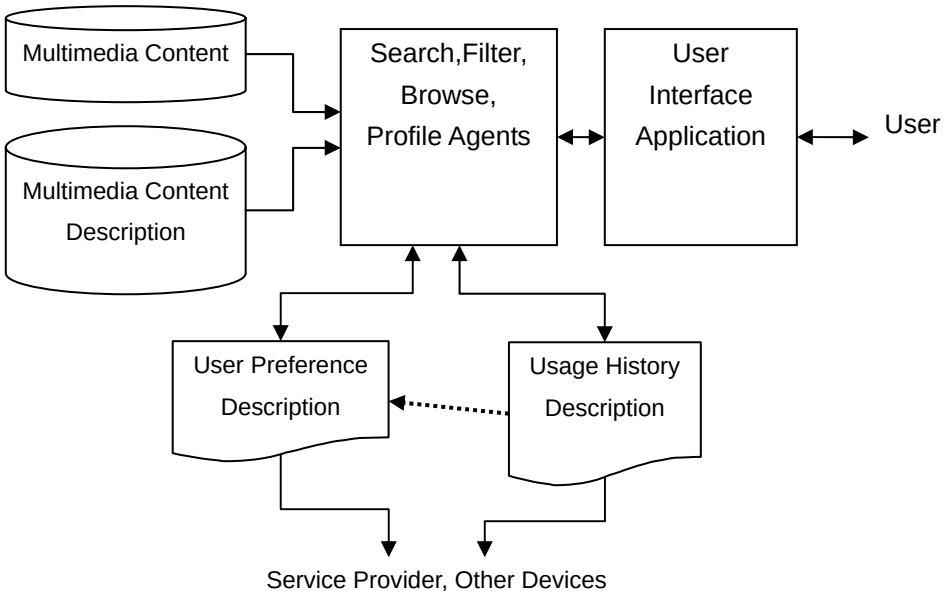

Figure 27. Overview of an interactive multimedia content personalization system

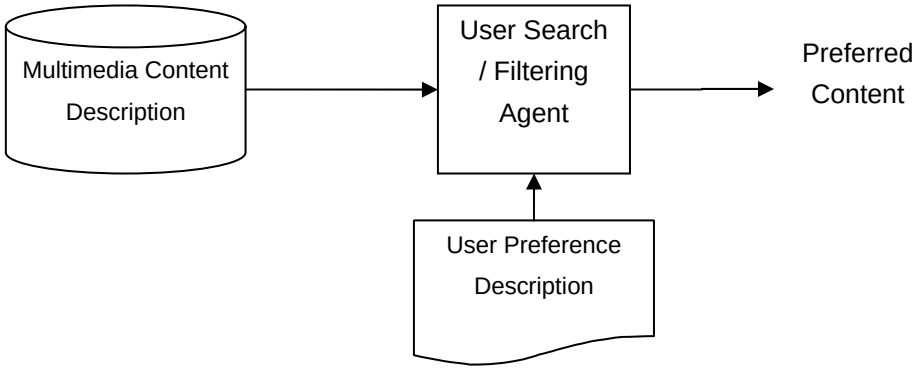

Figure 28. Overview of usage model for user Preference and content descriptions.

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.

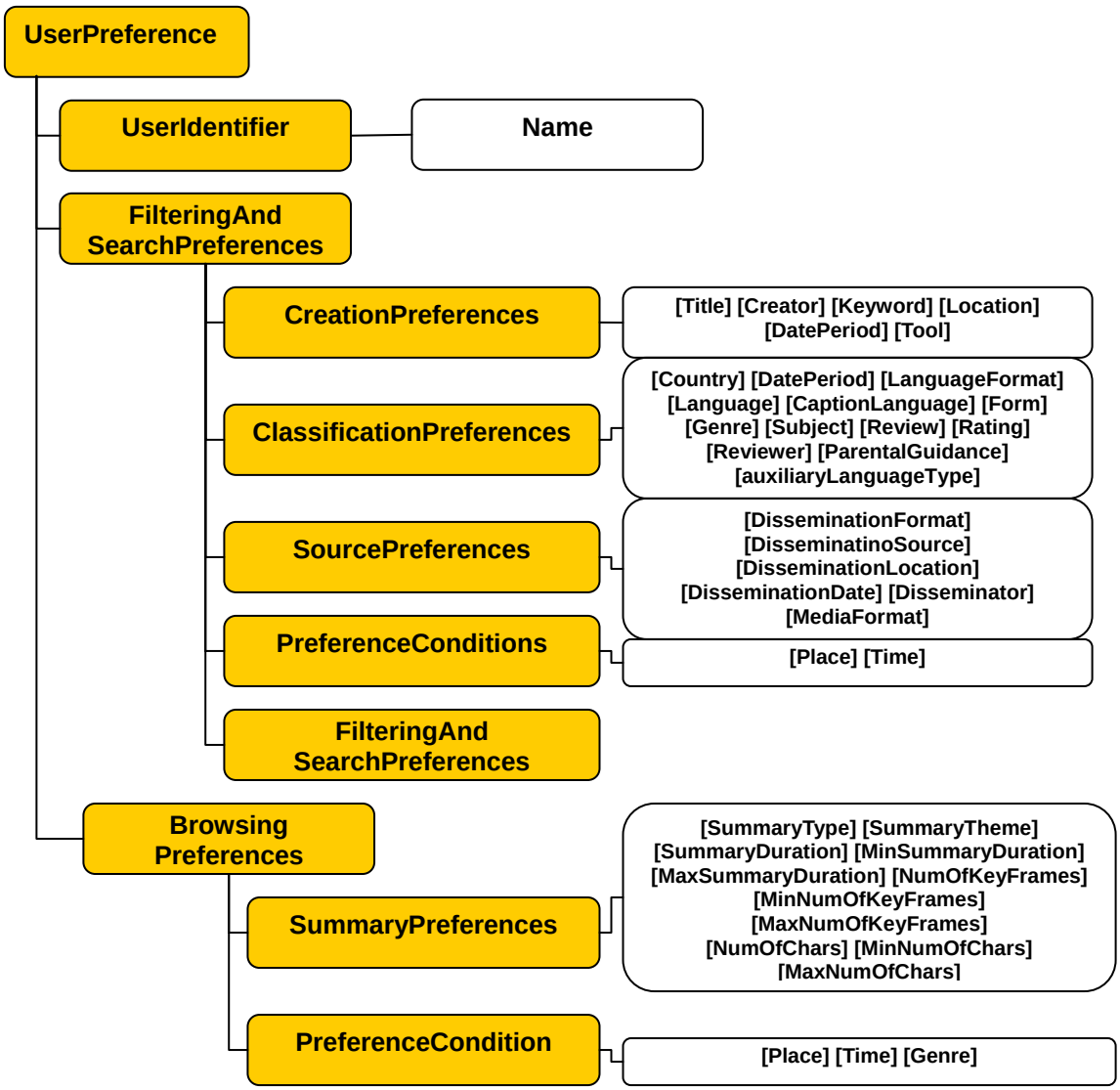

Figure 29. User Preference Description Scheme

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.

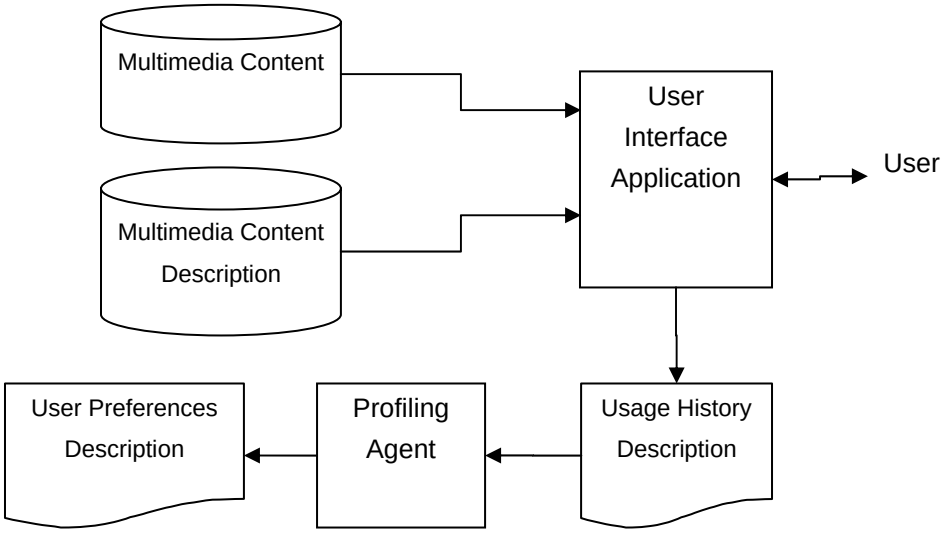

Figure 30. Overview of usage history descriptions and their applications.

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.

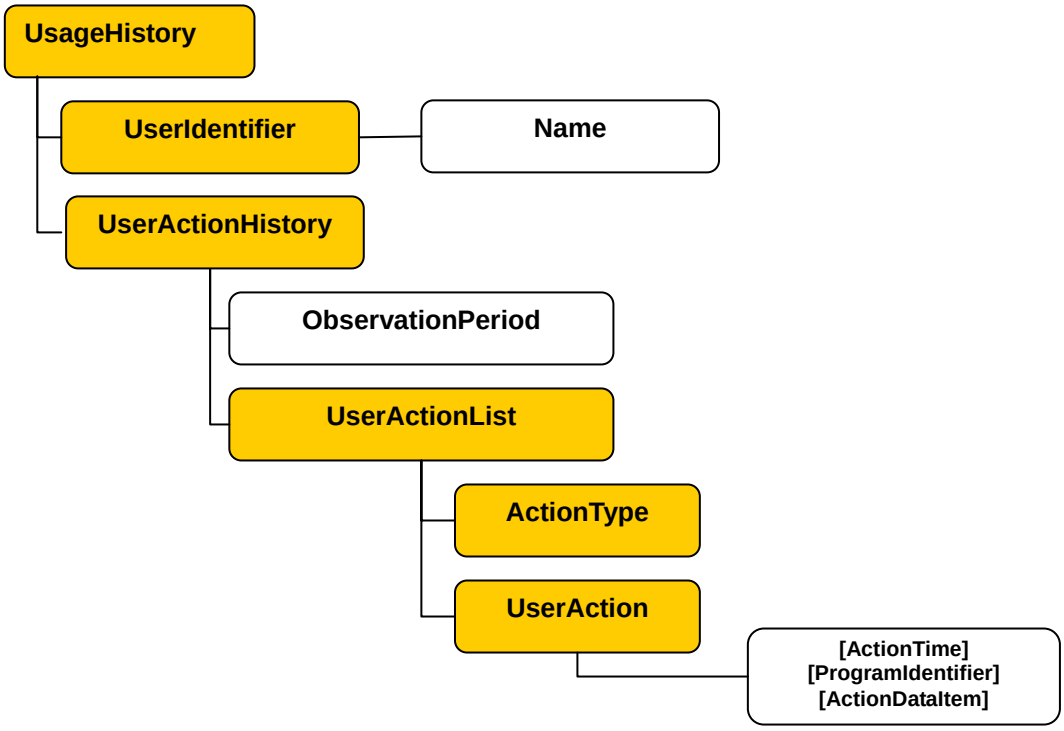

Figure 31. Usage History Description Scheme

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.

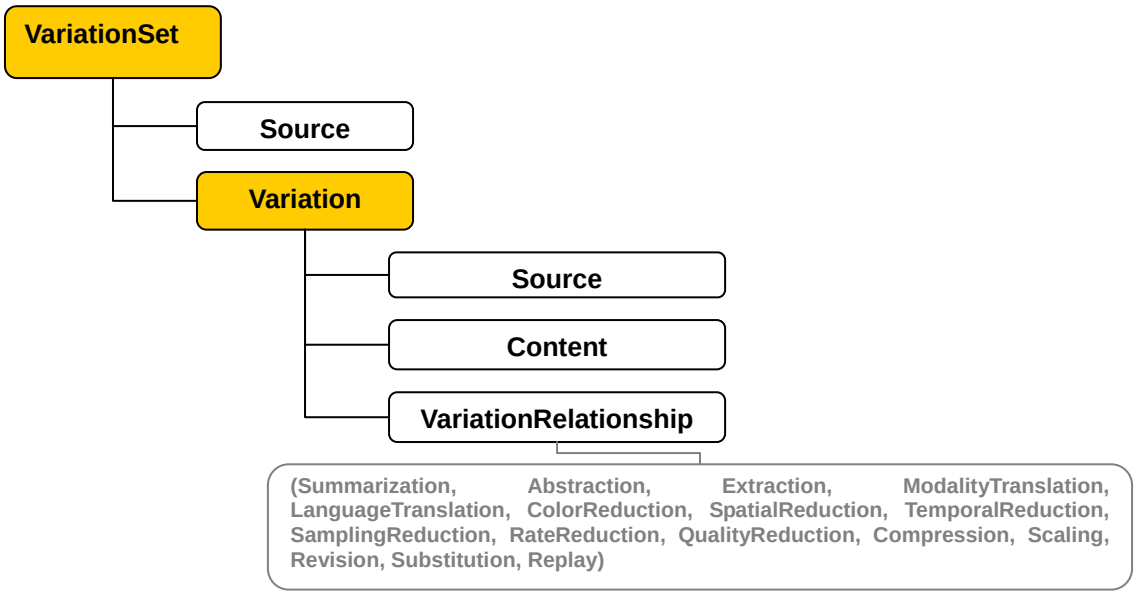

Figure 32. Variation Description

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. 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.

4.3.5. 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.

4.3.6. 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. 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.

5.1. 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 [16] 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].

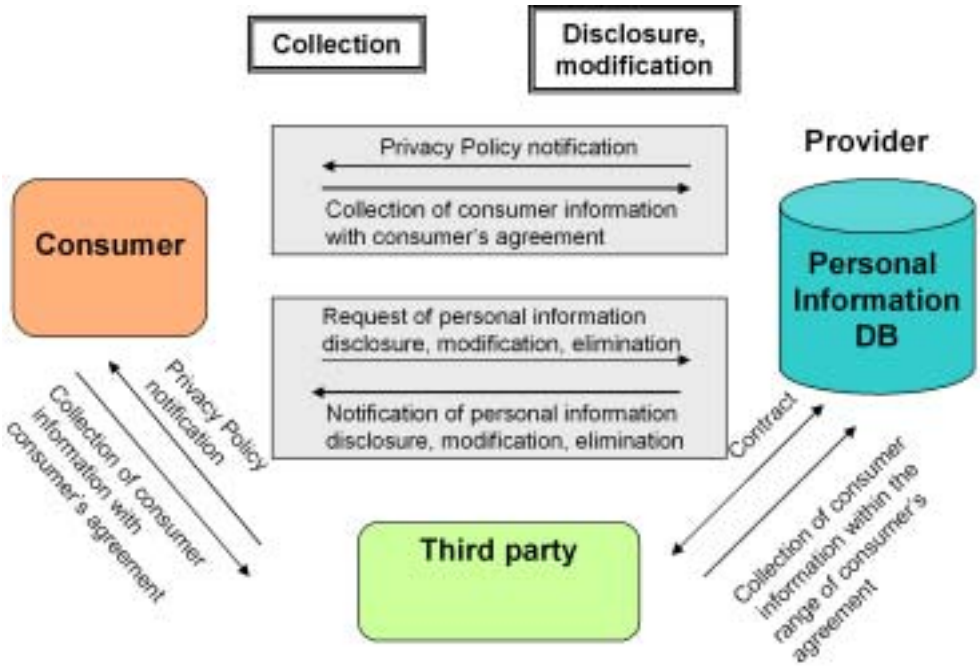

Figure 33. Framework of personal information protection

The term **'information privacy'** refers to the combination of communications privacy and data privacy [16]. 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 [17]. Privacy protection could be defined as a process of finding appropriate balances between privacy and multiple competing interests [16].

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

5.2. 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 5.5.2).

- **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.

5.3. 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.

5.4. 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 [13][14][15], 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.

5.5. Standards relevant to privacy

5.5.1. TRUSTe

TRUSTe **[19]** 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

5.5.2. P3P (Platform for Privacy Preferences 1.0)

The Platform for Privacy Preferences Project (P3P) **[20]** 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.

5.5.3. APPEL (A Privacy Preference Exchange Language)

APPEL 1.0 **[21]** 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.

5.5.4. MPEG-21 Part 5 and MPEG-21 Part 6

MPEG sees a **Rights Data Dictionary [22]** 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 [23]** 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, efficient creation of content metadata and description for adaptation, privacy protection, killer application, system designing, 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

The most important thing in UMA services is to adapt and deliver the contents in a way that maximizes the user satisfaction. There are three important issues in user-centric adaptation. 1) How to transform or select the content(s) in a way that provides the best experience to the user, 2) how to present the transformed or selected content(s) to the user to maximize their satisfaction and 3) how to measure the “quality of experience”.

1) Metrics for multimedia content transformation and selection.

The first challenge is how to transform or select the content(s) in a way that provides the best experience to the user. The problems are the following;

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

2) Presentation of the transformed/selected multimedia contents (and their metadata).

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.

3) Quality metrics.

The other challenge is how to measure the quality of experience. The evaluation metrics 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, 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.

- How to evaluate the quality of experience?
- 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.

6.2. 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.

- 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.

- 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. Privacy protection

As discussed in section 5, the main problems on privacy in UMA systems are as follows;

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.
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.
6) How to reconcile personalization with privacy.

6.4. UMA killer 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.

6.4.1. Examples of UMA Application

For more specific UMA application, it is important to make clear “what” the application delivers to “whom”. During the investigation on UMA, several applications considering my difficulties in my 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.

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 my pace. This system can be generalized not only for French studies, but to any multimedia e-learning contents.

Table 3 indicates the requirements for the universal multimedia e-learning system from the viewpoint of both user and service provider.

Table 3. 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.

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 4 indicates the requirements for the system from the viewpoint of both user and service provider.

Table 4. 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.

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.

6.5. System design

As all of the aspects described in 6.1 - 6.4 should be considered for UMA systems, especially in the adaptation process, 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.

1. Identification and selection of necessary descriptions.
2. How to locate of engines, descriptions and contents.
    - Where to locate UMA Engine. (consumer device/intermediate/provider side?)
    - Where to store personal data. (consumer device/intermediate/provider side?)
    - Where to locate usage environment tools.
    - Where to locate content descriptions.
3. System design to increase user awareness of what the system environments are doing, and how user information is collected and used and to empower users to manage their personal data.

6.6. Digital rights management and content usage control

Another relevant big challenge is digital rights management and content usage control.

- Restrictions on variation quality. (e.g. Commercial Film not allowed to be sent with low quality image)
- Restrictions on the length and context of the content to be sent. (e.g. some (part of) the content cannot be sent to some people of some category. Not to deliver a climax scene as a movie preview.)
- Definitions of the rights holders for derived variations. (e.g. the transcoded/transmoded variation belongs to whom?)

7. Conclusion

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

There are five key issues in UMA; User-centric multimedia content adaptation, (standardized) description necessary for adaptation, privacy protection, rights managements, and UMA application including its system designing.

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 transform or select the content(s) in a way that provides the best experience to the user, 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. Some examples were reported in this paper. 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.

References

[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).

[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).

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

[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.

[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.

[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.

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

[CC/PP]
Composite Capabilities/Preference Profiles. http://www.w3.org/Mobile/CCPP/

[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/

[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

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

References of Privacy

- Trust:
  [1] C. Nodder (Microsoft), "**Say versus Do; building a trust framework through users' actions, not their words.**", UbiComp 2003
  [2] M. Langheinrich (ETHZ), "**When Trust Does Not Compute – The Role of Trust in Ubiquitous Computing**", UbiComp 2003

- Location:
  [3] J.I.Hong (UCB), "**Privacy and Security in the Location-enhanced World Wide Web**", UbiComp 2003
  [4] G.Myles (Arizona Univ.), "**Preserving Privacy in Environments with Location-based Applications**", IEEE Pervasive computing, 2003.

- System design, general framework:
  [5] G.Karjoth and M.Schunter (IBM), "**A Privacy Policy Model for Enterprises**", 15th IEEE Computer Security Foundations Workshop (CSFW'02), 2002.
  [6] M. Langheinrich (ETHZ), "**Privacy Invasions in Ubiquitous Computing Environments**", UbiComp 2002.
  [7] 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).
  [8] M. Langheinrich (ETHZ), "**A Privacy Awareness System for Ubiquitous Computing Environments**", UbiComp 2002.
  [9] M. Ackerman (UCB), "**Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences**", ACM Conf. E-Commerce, 1999.
  [10] R. Bellotti, "**Design for Privacy in Ubiquitous Computing Environments**", ECSCW93, 1993.
  [11] CDT's guide to online privacy, http://www.cdt.org/privacy/guide/introduction/

- User behavior analysis:
  [12] R. Beckwith (Intel), "**Designing for Ubiquity: The Perception of Privacy**", IEEE Pervasive computing, 2003

- Capability Configuration, harmonization between CC/PP and P3P:
  [13] P. Osbakk and N.Ryan. "**Context, CC/PP, and P3P**", UbiComp 2002 Adjunct Proceedings, pages 9-10, Göteborg, Sweden, September 2002.
  [14] 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,
  [15] 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:
  [16] R. Clarke, "**Introduction to Dataveillance and Information Privacy, and Definitions**

of Terms", http://www.anu.edu.au/people/Roger.Clarke/DV/Intro.html, 1997.

[17] 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:

[18] Lorrie Faith Cranor, "**Web Privacy with P3P**", O'Reilly & Associates, ISBN 0-59600-371-4, 2002.

[19] "**The TRUSTe White Paper**", http://www.truste.org/

[20] "**The Platform for Privacy Preferences 1.0 (P3P1.0) Specification**", W3C Recommendation 16 April 2002, http://www.w3.org/TR/P3P/

[21] "**A P3P Preference Exchange Language 1.0 (APPEL1.0)**", W3C Working Draft 15 April 2002, http://www.w3.org/TR/P3P-preferences/

[22] "**MPEG-21 Part 5 - Rights Data Dictionary**", 2003.

[23] "**MPEG-21 Part 6 - Rights Expression Language**", 2003.

- Some active communities:
  - UbiComp http://ubicomp.org/
  - IEEE Pervasive Computing http://www.computer.org/pervasive/

Annotated References of Universal Multimedia Access

[Bormans03]
MPEG-21: The 21st century multimedia framework
Bormans, J.; Gelissen, J.; Perkis, A.;
Signal Processing Magazine, IEEE , Volume: 20 Issue: 2 , March 2003
Page(s): 53 -62

Abstract:
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.

- Content:

Context and motivation, Digital Rights Management, Heterogeneous Terminals and Networks, MPEG-21 Vision

- Specifications

  MPEG-21 once over lightly, Current MPEG-21 Parts

- MPEG-21 in a UMA Context

  UMA Context, Digital Item Adaptation, Example Use Case: Streaming Media, End-User Terminal

- Related Work

  Mediacom 2004, Capability exchange, DVB-MHP

- What's next?

[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.

[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]
Resource-driven MPEG-4 FGS for universal multimedia access
Chen, R.Y.; van der Schaar, M.;
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.

Contents:
- Introduction (device capabilities)
- Rate - Distortion - Complexity Framework
- Non-scalable MPEG decoding
- MPEG-4 FGS system
- Resource-Driven Streaming System

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

Page(s): 270 -271

Abstract: (almost the same content as [UMA-8].)
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.

Contents:
- Complexity profiling of MPEG-4 FGS decoding
- System level complexity scalability for FGS (apply different IDCT algorithms based on block type)
- Receiver-driven FGS streaming system

[Chung03]
"FGS-based video streaming test-bed for MPEG 21 universal multimedia access with digital item adaptation"
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;
Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on , Volume: 2 , 25-28 May 2003
Page(s): II-364 -II-367 vol.2

Abstract:
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

[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.

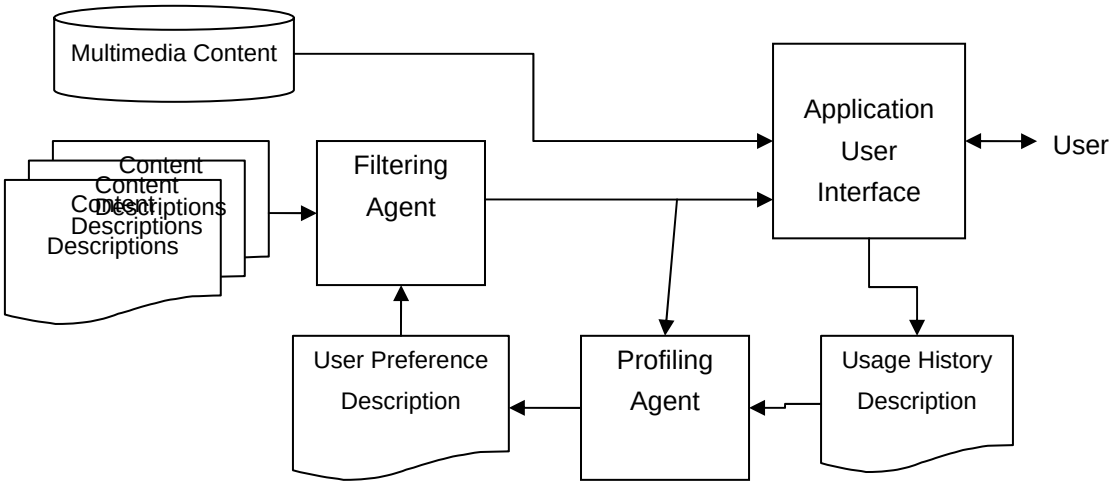

Figure A1. Content personalization framework

[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.

[Fossbakk01]

An MPEG-21 framework for streaming media

Fossbakk, E.; Manzanares, P.; Yago, J.L.; Perkis, A.;

Multimedia Signal Processing, 2001 IEEE Fourth Workshop on , 3-5 Oct. 2001

Page(s): 147 -152

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

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

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).

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

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

[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.

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

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

[Pereira03]
Universal multimedia experiences for tomorrow
Pereira, F.; Burnett, I.;
Signal Processing Magazine, IEEE , Volume: 20 Issue: 2 , March 2003
Page(s): 63 -73

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

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

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

[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.

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

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

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

[Vetro03]
From the guest editors - Universal multimedia access
Vetro, A.; Christopoulos, C.; Ebrahimi, T.;

Signal Processing Magazine, IEEE , Volume: 20 Issue: 2 , March 2003
Page(s): 16 -16

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.