Enhanced descriptions for personalized retrieval and automatic adaptation of audiovisual content

Iván Cantador, Fernando López, Jesús Bescós, Pablo Castells, and José M. Martínez
Escuela Politécnica Superior
Universidad Autónoma de Madrid

1 Introduction

The continuous growth of audiovisual content in digital format available worldwide poses new challenges to content retrieval technologies, calling for new solutions that cope with the current scale and complexity of content corpora, technological infrastructures, and user audiences. Two major problem areas to be addressed in facing such challenges involve a) the effective selection of content items when the scale of the retrieval space surpasses the capacity of users to query the corpus, or even browse search result lists; and b) the automatic adjustment of multimedia content to fit a wide variety of support infrastructures (terminals, networks, codecs, players, etc.), while making the most of the available delivery channel capabilities.

Addressing such problems implies work at the level of the identification, representation, dynamic analysis, and effective introduction of the contextual conditions that are relevant for the content retrieval and delivery processes, in order to best suit the situation at hand, in a way that optimizes the effectiveness in terms of user satisfaction. This involves building up a system awareness of dynamic conditions such as user interests and preferences of different kinds (high-level and low-level, broad and specific, explicit and implicit, content-related, source-related, etc.), device and network capabilities (screen resolution, network bandwidth, etc.). A proper description of multimedia content itself is needed, ranging from signal-level (colour, bitrate, etc.) to syntactic (objects, motion, visual attention, etc.) and semantic-level descriptions (topics, domain concepts, events, semantic relations).

This chapter focuses on a set of initiatives and achievements addressing such problems, resulting in different forms of personalized retrieval and dynamic adaptation (previous or simultaneously to delivery, not covered here) of multimedia content. The comprehensive view on multimedia adaptation provided here comprises low to high-level adaptation methods from the ranking of content units according to background user interests in different scenarios (e.g. presence vs. absence of an explicit user query, single vs. multiple users, etc.) to media adaptation techniques to different usage scenarios (terminals, networks, user preferences, etc.).

The chapter is structured as follows. The next section addresses the different aspects of content that need to be explicitly modelled in order to enable the adaptation techniques

discussed thereafter. Section 3 focuses on the personalization of content retrieval, the different sides and problems involved, and describes specific proposals in a semantic-based approach. After this, section 4 addresses the adaptation of content at signal-level, that is, customizing it in a content-agnostic way to the different terminals, networks and user preferences (from the media presentation point of view). Finally, some concluding remarks are provided in section 5.

2   Content modelling

Content access and delivery involve different processes or phases, such as query, selection (filtering), linking (recommendation), semantic adaptation, and signal-level adaptation (these last two steps could be combined in one), each involving different technologies. The automatic customization problem can thus be addressed at the different phases, motivating diverse requirements and approaches for each step of the chain. A common key aspect to all of them is the need for appropriate descriptions of the content to which the adaptation strategies apply, at the proper level and providing the relevant details needed to handle the content pieces, as required by the adaptive operations. Considering the specifics of multimedia content, the following aspects need to be taken into account to this respect:

- Representation model: the syntax and semantics associated to the bitstream that represents the signal used to deliver and present the content.

- Description model: the syntax and semantics associated to what is present in the content from a signal-level point of view (e.g. colour, shape, texture, motion for visual content) and from a mid-level point of view (generic objects and events), as well as the descriptive information required from an archival point of view.

- High-level semantics model: the syntax and semantics associated to the high-level interpretation and meaning of what is present in the content, generally linked to a particular application domain.

The requirements and available technologies related to each of the above description levels are discussed in the next sections.

2.1   Content representation

There are different specifications for the structuring of the bitstream used to represent media content. Although proprietary formats were in use in the past, nowadays most applications work with standardized formats defining both the syntax and the semantics of the bitstream, from a purely signal point of view, and mainly describing a compressed representation of the media, with a good subjective quality even in the presence of heavy compression.

Most of the wide-spread standards for audiovisual content belong to the JPEG family for images (the well-known JPEG format available in all digital cameras, and the new

JPEG2000 providing much better quality for the same bitrate), and to the MPEG family for video and audio: from MPEG-2 used in DVDs and Digital Television, and MP3 for audio on the Internet, to heavily improved versions in the form of MPEG-4 Advanced Audio Coding (AAC) and MPEG-4 Advanced Video Coding (AVC) also known as H.264.

2.2   Content description

Besides the content representation of the signal, it is required to have a description of the content, that is, metadata (data about the content essence). These descriptions range from the classical archival data (title, author, release date, abstract, keywords, etc.) to content-based descriptions of what is present in the content (e.g., a goal in a soccer video, nudity in a film, an interview by the anchorman in a news program). There are several standards for archival metadata (e.g., EXIF for images, ID3 for audio files, Dublin Core for generic digital content), whilst only a small number of standards, mostly broadcasters-oriented, deal to some extent with content-based metadata (e.g., SMTPE Metadata Dictionary, EBU P/Meta, TV-Anytime).

The MPEG-7 standard [17] is the most comprehensive multimedia content description specification, covering both archival metadata as well a large number of metadata for content-based description. MPEG-7 provides description tools for low-level audiovisual descriptions (e.g. colour, shape, audio spectrum), mid-level (e.g. face description, musical timbre), structural level (e.g. shot and scene decomposition, musical movements), and even tools for semantic and linguistic description. The main drawback of the MPEG-7 specifications is their complexity, mainly due to the fact that MPEG-7 was aimed to be a generic standard, not focused on a particular application domain. In order to help reducing this complexity, and following previous MPEG standards, profiles (subset of tools) have been defined for different applications domains.

2.3   Usage Context description

By Usage Context we understand the terminal capabilities, network conditions and media related user preferences (e.g. preference for an image slide rather than a video, or a travelling video instead of an image that doesn't fit in the screen) that are active in each media consumption session. All this information is required in order to make decisions about what kind of adaptation should be performed over a media for a specific session. There are several standards dealing with some of these descriptions (e.g., UAProf, CC/PP), but the Digital Item Adaptation (DIA) part of MPEG-21 [31][4] is the one covering the more complete set of such descriptions, as well as additional tools for media adaptation techniques.

2.4   High-level semantics

The descriptions discussed above convey a lot of useful information for handling and reformatting content at the signal-level, as described later in section 4. However, from the point of view of cognitive user needs, higher-level information is required. At the end of the consumption chain, users of multimedia content are concerned with the information they are going to get by viewing a specific document, which is a major source of relevant

input for personalizing the choice of documents to be delivered, or fragments to be viewed within. High-level semantics aim to describe what objects appear in a scene (people, cars, buildings, trees, roads), what type of scene is displayed (tennis, beach, restaurant, indoors/outdoors), what happens in it (a person enters the room, walks, sits down, smiles, eats), and so forth.

The type of entities, events, and subjects that may appear in a multimedia document are virtually *anything*, which makes it hard to provide a general framework supporting the needed descriptions in a general way. While it is indeed impossible to model the world as a whole (although some attempts have been made [13]), a partial and feasible approach is possible for restricted domains, as supported by ontology-based knowledge technologies [24]. Ontology-driven representations, and in particular the ones supported by W3C standards such as RDFS [3] and OWL [19], have nice properties such as being formal, non-ambiguous, rich (in direct proportion to the human effort invested), and enabling automatic inference based on Description Logics [2].

In the techniques discussed in the next section, the proposed representation to describe the meanings within content consists of a list of domain concepts that appear or happen in the content [7]. Concepts are associated to content by semantic annotation links, which may include time stamps when concepts occur in specific content segments, and one or several weight values indicating the strength of the link. The weights can reflect the importance of the concept in the content, the certainty that the concept actually occurs (e.g. when it has been recognized by an automatic content analysis technique), or any other relevant numeric measures. Ontology-based annotations can be created manually, or obtained through automatic means [9], e.g. by extracting concepts from manual textual annotations, spotting words or known sounds in audio tracks, detecting objects and events in visual scenes, or by coordinating several of these approaches.

This representation goes beyond the simpler and currently dominating forms of free annotation, commonly consisting of plain string keywords or arbitrary text sentences. It provides a unified, unambiguous representation of the semantic space for annotation, hence a solid ground upon which powerful adaptation techniques can be devised, capable of making sense of the high-level semantic descriptions. Additionally, all the aforementioned facilities supported by ontology-based technologies are available to the advantage of the development of personalization strategies.

3   Personalized content retrieval

In general terms, personalizing the selection of content for user access involves knowing something about the user beyond her last single request, and taking advantage of this knowledge in order to improve the system response to the actual user needs, in terms of the utility of the retrieved content for each individual user. Room for such improvement exists increasingly often, to varying degrees, in common retrieval scenarios, either because the request is vague, in the sense that too much content matches it for the user to process (e.g. as happens most of the time in large-scale content spaces), or because there is no explicit request (e.g. the user is not instantly aware of new content of interest). In addition to the requirements in the scope of content description, the problems to be ad-

dressed to realize this view can be related to the areas of representation, elicitation, and exploitation of user preferences.

3.1   User preference modelling

Modelling user preferences is central to content personalization technologies, as it forms the basis on which the system may automate decisions on behalf of users. User profiles for multimedia retrieval provide a logical view of known or predicted user particularities and needs with respect to content features and formats, in a suitable form for use by the algorithms that implement the automatic adaptation strategies. A variety of structures and paradigms have been used in research and industry for the representation of relevant user information in this context. Among standardization efforts, the MPEG-21 DIA stands out as one of the most exhaustive specifications covering, among other more media-related features, information such as demographic data, general preferences, audiovisual presentation preferences, disabilities, focus of attention, etc. The Multimedia Description Schemes in MPEG-7 provides additional structures for storing usage history data [17]. MPEG-21 describes the user information which, along with content descriptions in MPEG-7, provides a suitable basis for multimedia-specific content adaptation in the delivery phase, as will be described in section 4.

Related to content selection, the usage environment data described in MPEG-21 is generally not enough (or not relevant) to make far-reaching predictions of what content choice a user might enjoy. To this respect, it is more important to know *what the content is about*, and what subjects, concepts or objects *the user is interested in*. A common approach to model this kind of preference is the vector-space model, in which user interests are represented as a vector of weights defining the intensity of interest for different *things*. The space of such things, in terms of which user interests are described, can correspond to elements of diverse nature, such as individual documents, categories from a taxonomy, or even plain words [8], [12], [20]. In our research, we have explored the potential of enhanced representations of meanings as the foundation of this preference space, beyond prior simple approaches, as a means to enable improvements in the reach and accuracy of personalization [5], [10]. More specifically, we have developed ontology-based preference models, where the preference space is based on domain ontologies [24], in such a way that user interests are represented by a weight for each domain concept in the ontology. These weights can be derived from the observation of user activity (queries and accessed content) over a period of time, and the repeated occurrence of concepts related to the objects involved in this activity. This is of course a problem on its own; the reader is referred to e.g. [5], [21] for further discussion.

The advantage of using ontologies as the representational grounding for the semantic space lies in the precise and detailed information that is made available for the system e.g. to match user preferences to content descriptions, enabling substantially more elaborate strategies than simpler representations support. For instance, the system can derive new user preferences from the initial ones based on the extra knowledge supplied by an ontology, using the formal inference mechanisms supported by ontology standards. This is further explained when we discuss the use for context modelling and content filtering in the next sections.

An alternative (or complementary) view on the personalized recommendation of content is the social one, in which users are not just analyzed in isolation, but in comparison to other users. A well-known approach in this direction is the so-called collaborative filtering strategy, which is based on the general assumption that users with common traits may enjoy the same or related content [1]. This principle raises the problem of measuring the similarity between users, which can be done based on a comparison of the profile information (e.g. semantic preferences, demographic information), or the history of common choices. The similarity of picked contents can be in turn measured by functions other than straight equality, where again, a rich semantic representation of content enables the detection of indirect similarities between items, that would not be found by checking for plain coincidence. For instance, a user who liked the films *Interiors* and *Annie Hall*, and a user who liked *Broadway Danny Rose* and *Manhattan* would be considered candidates for a possible affinity based on indirect evidence, since they both enjoyed films directed by *Woody Allen*, even if they did not watch the same films. This is possible if a domain ontology on cinema is available to the system, where movies, directors, actors, etc., are defined as interrelated domain concepts and individuals.

3.2   Context modelling

In order to properly customize content, knowledge about the context in which the content is sought and consumed by a particular user at a given time, is relevant as well. Research in this area has commonly considered user preferences and terminal and network capabilities (which we discuss later in this chapter) as part of the context model. Beyond this, other elements addressed in the literature include the rendering context (e.g. noise, illumination), location, time, meteorological data, etc.

Besides the user preference information described in the previous section, in our research on content retrieval adaptation we consider three additional contextual dimensions, namely a) the dynamic user focus, consisting of a weighted, semantically coherent set of domain concepts that have been involved, in some way or other, in an ongoing user session [28]; b) the semantic context of meanings, defined as the domain concepts closely related (through semantic relations explicitly defined in an ontology) to a given set of concepts [21]; and c) the social context of a user, consisting of the sets of users related to her in different possible ways [6].

The semantic context of a concept is given by the concepts around it in the semantic network defined by a domain ontology, based on the relations that interlink the concepts. Semantic paths provide a basis to define distance measures between concepts, upon which we build fuzzy contextual supersets for a given set of concepts [28]. This step is key in our approach to handle the social context, in which further similarities between users are found by comparing the semantic context of their preferences [6]. This is applied in a similar way in order to take advantage of the live user focus to contextualize persistent user preferences, and enable a more accurate personalization of retrieval results [21], as will be described in the next section.

3.3   Content filtering

Based on the relevant knowledge about user interests that a system is able to capture and elaborate on, different techniques can be used to personalize search and retrieval results, which come into play at different points of a retrieval system. For instance, user preferences have been used to reformulate user queries, by expanding or refining the information in the query (e.g. adding, reweighing, or even disambiguating terms), using information from the user profile and history [23]. Preferences can also be applied on the search result after the query has been answered. For example, the results can be re-ranked by a complementary similarity measure between documents and user preferences [8], [20]; clustering techniques have also been proposed to group results in sets of categories, ranking higher the more relevant categories by user preference [14]. A long series of variations of the popular PageRank algorithm for Web search have been researched as well, which use different initialization strategies based on user preferences, in such a way that the resulting PageRank value is biased towards individual user interests [12]. In scenarios where information (preferences, history) of a large number of users is centrally available to the personalization system, it is also possible to apply collaborative filtering techniques in a way that thousands of users benefit from each other's experience without even getting ever acquainted, as discussed earlier.

We have explored the potential for improvement that can be gained in such different strategies by building on the semantic enhancements for the representation of user interests described in section 3.1. To begin with, ontologies enable the automatic extension of user preference reach by means of inference steps. For instance, a preference for pets would automatically expand to cats and dogs. On a more complex path, a preference for award-winning films can be defined in an ontology language like OWL by a restriction class defined as the movies that have won *some* award. The user can have his preference fulfilled by the system inferring that, say, an Oscar is a subclass of award, and a movie that won *some* Oscar is an instance of the preferred class by the user.

Our primary application of these principles has been brought to a basic content filtering approach, which re-ranks search results by comparing user preferences and content metadata in the ontology-based vector space as described in earlier sections. This comparison is achieved by a similarity measure that compares these two concept vectors (user preference and content semantics) based on the cosine function. This measure is then combined with the relevance score returned by the search engine (which is taken here as a black box) for each content item and the query being answered, thus introducing a personalized bias in the ranking [5].

This approach is further elaborated by contextualizing the preference model before it is matched to content descriptions, based on the live user focus introduced in the previous section. In our approach, putting preferences in context means to activate only those persistent preferences that are semantically close to what is going on in the session and what the user has in mind at that time. This matches the way human preferences work in real life where, for instance, bringing up a long-term (e.g. professional) user interest for pharmaceuticals when a user is searching for a good movie in the weekend could be generally out of place.

This feature is addressed as follows. During a retrieval session, the system collects all the concepts involved in user queries and metadata of the content selected by the user. A

vector of weighted concepts (where weights decay with time) is built in this way, which is taken to be representative of the short-term user attention, or at least to be related to it. Then, the contextualization mechanism starts by estimating the semantic distance between each concept in the preference set and the live focus set, which is achieved by scanning the number and length of paths connecting them through the semantic network defined in the ontology. This is achieved by computing the extended fuzzy semantic context of user preferences and focus, and their fuzzy intersection. The resulting degrees of membership of concepts to this fuzzy set is combined with the original preference weight in the long-term user profile to produce a contextualized preference vector, where the weight of preferences that are semantically unrelated or far from the current user focus will be close to zero, thus achieving the desired effect [21], [28].

The semantic context modelling and user profile extension techniques have also been applied in our research to elaborate on the collaborative recommendation approach, as introduced in section 3.1. Besides the extension of similarity functions based on indirect semantic comparisons discussed earlier in this chapter, we have explored a step beyond this by further refining the model with the notion of semantic layers within user preferences, as we explain next. In typical collaborative filtering approaches, the comparison between users is performed globally, in such a way that partial, but strong and useful similarities might be missed. For instance, two people may have a highly coincident taste in cinema, but a very divergent one in sports. Their opinions on movies could be highly valuable for each other, but risk to be ignored by many recommender systems, because the global similarity between the users might be low. We thus contend for the distinction of different layers within the interests of users, as a useful refinement to produce better recommendations.

This idea is achieved by analyzing the structure of the domain ontology, the weighted links between users and ontology concepts (as defined by preferences), the links between concepts and contents (annotations), and the links (explicit ratings) between content and users. Based on this rich interrelation within and across the three spaces (users, concepts, content), we develop strategies of coordinated clustering which produce focused recommendations based on partial but cohesive similarities. Our approach finds groups of interests shared by users, and communities of interest among users. Users who share interests of a specific concept cluster are connected in the corresponding community, where their preference weights determine the degree of membership to that cluster. This enables focused recommendations in the different communities [6]. At the same time, this approach tackles the common sparsity problems of the collaborative filtering approach in large-scale retrieval spaces (i.e. user preference information being scarce when an item is new, a user is new, or user ratings are thinly spread over a huge number of items [1]), by finding indirect similarities among users and content items. Note that links between users (contacts) and between documents (e.g. hyperlinks or other cross-references) could also be considered as part of the available information to further enhance the analysis, which we foresee as future research.

4   Content adaptation

Once a specific content item is picked by the user from a list of choices retrieved by the system according to user queries, preference, or any of the strategies discussed in the previous section, the time comes to actually deliver the content itself in the most suitable form. At this point, the development of new access networks providing multimedia capabilities, and a wide and growing range of terminals, makes the adaptation of content a major issue for future multimedia services. Content adaptation is the main objective of a set of technologies that can be grouped under the umbrella of Universal Multimedia Access (UMA) [29]. This means the capability of accessing to rich multimedia content through any client terminal and network. In this way content adaptation bridges content authors and content consumers in a world of increasing multimedia diversity.

In order to perform content adaptation it is necessary to have the description of the content and the description of the usage environment. To enhance the user's experience [22], not only terminal and network characteristics and conditions should be taken into account when adapting, but also user preferences and disabilities, as well as environmental conditions. All this information imposes some constraints to the content to be obtained after adaptation.

4.1   Content adaptation taxonomy

According to *the level of understanding applied to the media*, multimedia adaptation can be performed in two different ways:

- **Signal level adaptation**, committed to transcoding media resources without knowledge of the meaning of the content.

- **Semantic level adaptation**, which modifies the media assuming that there exists some knowledge about the meaning of the content.

According to *the information used to decide the adaptation to perform*, multimedia adaptation can be based on three different sources:

- **Usage environment driven adaptation**. The terminal, network and user preferences are taken into account to decide the adaptation to perform to the multimedia content.

- **Semantic driven adaptation**. Evaluates the content of the media to select the more relevant parts. This kind of adaptation has been mainly used to perform content summarization.

- **Perception driven adaptation**. Performs transformations according to some user preferred sensation, or assists a user with a certain perception limitation or condition. The user perception of the content will be different from its original version, for example, to address the needs of users with visual handicaps (e.g. colour blind deficiencies, or specific preferences) in terms of visual temperature sensations.

This class of customization operations considers all types of adaptations related to some human perceptual preferences, characteristics or handicaps.

According to *the way to group media*, adaptation can be performed at two different levels:

- **Media level adaptation,** committed to adapt a unique resource, named **media resource**, and usually stored as a file. Content to be adapted at this level can also be represented as a stream of bytes. Typically, a standardization body defines the whole format of this single modality media

- **System level adaptation** (also named **multimedia level**), committed to adapt one or more resources grouped as a compound resource, named **system resource** (e.g. a web page, or a SMIL file). This system resource can also convey metadata (descriptors). Within the system level compositions, we can identify three main kinds of adaptation: 1) structure level adaptation adapts a resource that is the union of (or has references to) other media resources (e.g. DI, HTML); 2) layout level adaptation changes the arrangements of the constituent elements in the scene (e.g. HTML, SMIL); 3) synchronization level adaptation modifies the timeline of the constituent resources (e.g. MPEG files summarization, or SMIL image serialization). For example, an MPEG-21 Digital Item Declaration [4] is capable of conveying media resources and descriptors grouped as a system resource named Digital Item (DI). During the adaptation, the number of resources and descriptor elements may change.

These two adaptation levels (media and system) can be performed in a signal or semantic way. Conversely, signal and semantic level adaptations can be performed in a media or system level.

Another dimension to consider in an adaptation framework are the different *adaptation approaches,* that can be also applied with content understanding (semantic level adaptation) or without it (blind or signal level adaptation); all these approaches are expected to coexist [30]:

- **Transcoding**. This is the most frequent adaptation, where the media parameters are changed in order to fit the constraints imposed by the usage environment. Scaling, colour reduction, coding standard, etc.

- **Transmoding**. This adaptation implies a change of modality of the original media in order to enable it to be presented in the selected scenario. Text to audio, video to slideshow, image to video, etc. are among this new generation of content adaptation.

- **Scalable content truncation**. Scalable formats are based on the co-existence in the same file or stream of different versions of the same content at different tempo-

ral, spatial and quality resolutions. Efficient tools for management of such content are provided by bitstream description tools within MPEG-21 DIA[31].

- **Summarization**. Video and audio skimming (creating excerpts of the temporal content –e.g. video trailers) are the most common approaches to summarization, although there are other proposals based on additional transmoding (e.g., a video poster consisting of the composition into one poster image of selected keyframes of the summarised video). The presentation of information provides different options to summarization, whilst the common ground are the content analysis technologies for the identification of the relevant parts of the content: keyframes or relevant video segments of a video, chorus of a song, main theme of a symphony, etc.

4.2   Content adaptation engine

Content adaptation is performed by a module of the complete content retrieval system, that is usually named Content Adaptation Engine. It aims to provide the user the best experience when accessing the requested content within the current usage environment. A Content Adaptation Engine includes two main functionalities: deciding the adaptation to perform, and managing the decided adaptation process.

In the simplest cases an Adaptation Engine is only able to perform one adaptation (e.g., reducing the spatial resolution of an image to a half in each dimension) and therefore no decision phase is required.

The next complexity step comes up when the Content Adaptation Engine can choose between different values for certain parameters of the adaptation (e.g. reducing the spatial resolution in a predefined range). Additional complexity appears when a Content Adaptation Engine is able to perform two (or more) adaptation processes, each with its own parameters (e.g. for achieving a target file size, spatial resolution or bits per pixel can be reduced). In this case, the decision is not only parameter-based; instead, a combination of adaptation processes and their parameters can achieve the target adaptation.

According to the identified functionality, a generic Content Adaptation Engine can be modelled to include two main submdules: a Decision Module and a set of Content Adaptation Tools (CATs). Each CAT is able to perform a particular adaptation. The different available CATs may diverge in the adaptation approach (e.g. transcoding, transmoding, etc.), the range of parameters values, the supported input and output formats, the performance (in terms of processing requirements, quality, etc.), and so forth. Hence, there is a need for describing the CAT adaptation capabilities [26], so that the Decision Module is able to incorporate also this description when making a decision primarily based on the aforementioned content and usage environment descriptions.

Several approaches have been proposed to perform content adaptation [25], [31], [16], [11], [18]. We have developed a Content Adaptation Engine, named CAIN (for Content Adaptation Integrator) [18] , aimed to the integration of different content adaptation approaches, that is, different CATs [27]. In CAIN the Decision Module uses Constraints Programming [32] for selecting the CAT that is best suited to the optional and mandatory constraints imposed, respectively, by terminal and network characteristics and user preferences [15]. It should be noted that [11] proposes the use of a scheduling algorithm to

find a chain of elementary adaptation operations which transform the media accordingly, whilst the CAIN framework considers CATs that perform several elementary adaptations. The Decision Module selects only one CAT from the available ones (we are evaluating to extend our solution in order to allow the concatenation of CATs in the future). The adaptation methods are constructed upon the foundations provided by MPEG description standards: content descriptions are based on MPEG-7 MDS [17] and MPEG-21 DIA BSD [4], whilst the context descriptions are based on a subset of MPEG-21 DIA Usage Environment Descriptions tools [4].

5   Conclusions

Automatic adaptation is a major issue in modern multimedia content delivery environments and infrastructures. In this chapter we have discussed a set of adaptation techniques that address this need, spanning across the whole content retrieval and delivery cycle, from the selection and choice of content units, based on high-level descriptions of the meanings within, to the actual delivery of content streams, adapted to the available conditions at the consumer end-point.

The main innovations in our proposed approaches for personalized retrieval focus on the potential for improvements enabled by working on a) the representation of semantics and b) the consideration of the retrieval user context, where in our model so far the latter includes semantic contexts, the social environment, and the user focus. The ontology-based approach to semantics representation has its own tradeoffs, the main ones being the limited availability of ontologies, and the development cost and formalization problems involved in defining very detailed ones. However, the proposed personalization techniques are tolerant to incomplete knowledge, which means that they make the most of high-quality semantic information whenever it is available, and they degrade gracefully to the performance of any standard technique based on simpler representations (e.g. keywords, documents) with which our techniques can be combined, as the completeness of ontological knowledge decreases for the domain area at hand.

The proposed techniques for content adaptation are characterized by the capability to both decide on the most appropriate adaptation strategy for the dynamic situation in process, and actually carry out and manage the selected adaptation approach. The adaptation of content is built upon the expressive capabilities of the MPEG standards. Content and access adaptation together address the challenges raised by the new order of magnitude in the scale of current content environments, and the heterogeneous and dynamic nature of delivery platforms and user audiences, in order to deliver the subjective content service quality that an enhanced user experience requires. In order to reach this point, automatic content understanding is currently the most challenging research issue in content analysis in order to be able to perform semantic based adaptation based on what is present in the content.

Acknowledgements

References

[1] G. Adomavicius and A. Tuzhilin. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE Transactions on Knowledge and Data Engineering 17(6), 2005, pp. 734-749.

[2] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider. The Description Logic Handbook: Theory, Implementation, Applications. Cambridge University Press, 2003.

[3] D. Brickley and R. V. Guha. RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation 10 February 2004.

[4] I. S. Burnett, F. Pereira, R. Van de Walle,and R. Koenen. The MPEG-21 Book. Wiley & Sons, 2006.

[5] I. Cantador, M. Fernández, D. Vallet, P. Castells, J. Picault, and M. Ribière. A Multi-Purpose Ontology-Based Approach for Personalised Content Filtering and Retrieval. In M. Wallace, M. Angelides, Ph. Mylonas (Eds.), Advances in Semantic Media Adaptation and Personalization. Springer Verlag Studies in Computational Intelligence, Vol. 93, February 2008.

[6] I. Cantador and P. Castells. Multilayered Semantic Social Network Modelling by Ontology-Based User Profiles Clustering: Application to Collaborative Filtering. 15th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2006). Podebrady, Czech Republic, October 2006. Springer Verlag LNCS Vol. 4248, 2006, pp. 334-349.

[7] P. Castells, M. Fernández, and D. Vallet. An Adaptation of the Vector-Space Model for Ontology-Based Information Retrieval. IEEE Transactions on Knowledge and Data Engineering 19(2), February 2007, pp. 261-272.

[8] P. A. Chirita, W. Nejdl, R. Paiu and C. Kohlschütter. Using ODP metadata to personalize search. 28th International ACM SIGIR Conf. on Research and Development in information Retrieval (SIGIR 2005). Salvador, Brazil, August 2005, pp. 178-185.

[9] S. Dasiopoulou, V. Mezaris, I. Kompatsiaris, V. K. Papastathis, and M. G. Strintzis, Knowledge-assisted semantic video object detection, IEEE Transactions on Circuits and Systems for Video 15(10), October 2005, pp. 1210-1224.

[10] S. Gauch, J. Chaffee, J., and Pretschner, A. Ontology-Based Personalized Search and Browsing. Web Intelligence and Agent Systems 1(3-4), April 2004, pp. 219-234.

[11] D. Jannach, K. Leopold, Ch. Timmerer, and H. Hellwagner. A knowledge-based framework for multimedia adaptation. International Journal on Applied Intelligence 24(2), April 2006, pp. 109-125.

[12] G. Jeh and J. Widom. Scaling Personalized Web Search 12th International World-Wide Web Conference (WWW 2003). Budapest, Hungary, May 2003, pp. 271-279.

[13] D. Lenat and R. V. Guha. Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project. Addison-Wesley, 1990.

[14] F. Liu, C. Yu, and W. Meng. Personalized Web Search For Improving Retrieval Effectiveness. IEEE Transactions on Knowledge and Data Engineering 16(1), January 2004, pp. 28-40.

[15] F. López, J. M. Martínez, and V. Valdés. Multimedia Content Adaptation within the CAIN framework via Constraints Satisfaction and Optimization 4th Int. Workshop on Adaptative Multimedia Retrieval (AMR 2006). Geneva, Switzerland, July 2006.

[16] J. Magalhaes and F. Pereira. Using MPEG standards for multimedia customization. Signal Processing: Image Communications 19(5), 2004, pp. 437-456.

[17] B. S. Manjunath, P. Salembier, and T. Sikora. Introduction to MPEG-7 Multimedia Content Description Interface. Wiley & Sons, 2003.

[18] J. M. Martínez, V. Valdés, J. Bescós, and L. Herranz. Introducing CAIN: A meta-data-driven content adaptation manager integrating heterogeneous content adaptation tools. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2005). Montreux, April 2005.

[19] D. L. McGuinness and F. van Harmelen. OWL Web Ontology Language Overview. W3C Recommendation 10 February 2004.

[20] A. Micarelli and F. Sciarrone. Anatomy and Empirical Evaluation of an Adaptive Web-Based Information Filtering System. User Modelling and User-Adapted Interaction 14(2-3), February 2004, pp. 159-200.

[21] Ph. Mylonas, D. Vallet, P. Castells, M. Fernández, and Y. Avrithis. Personalized information retrieval based on context and ontological knowledge. Special Issue on Contexts and Ontologies, Knowledge Engineering Review 23(1), March 2008.

[22] F. Pereira and I. Burnett. Universal Multimedia Experiences for Tomorrow. IEEE Signal Processing Magazine 20(2), March 2003, pp. 63-73.

[23] J. Rocchio. Relevance feedback information retrieval. In Salton, G. (Ed.), The SMART Retrieval System: Experiments in Automatic Document Processing. Prentice-Hall, 1971, pp. 313-323.

[24] S. Staab, and R. Studer (Eds.). Handbook on Ontologies. Springer Verlag, Berlin Heidelberg New York, 2004.

[25] B. L. Tseng, C. Y. Lin, and J. R. Smith. Using MPEG-7 and MPEG-21 for Personalizing Video. IEEE Multimedia 11(1), March 2004, pp. 42-53.

[26] V. Valdés and J. M. Martínez. Content Adaptation Capabilities Description Tool for Supporting Extensibility in the CAIN Framework. International Workshop on Multimedia Content Representation, Classification and Security (IWMCRS 2006). Istambul, Turkey, September 2006.

[27] V. Valdés and J. M. Martínez. Content Adaptation Tools in the CAIN framework. In L. Atzori, D. D. Giusto, R. Leonardi, F. Pereira (Eds.), Visual Content Processing and Representation. Sprimger Verlag LNCS Vol. 3893, 2006, pp. 9-15.

[28] D. Vallet, P. Castells, M. Fernández, P. Mylonas, and Y. Avrithis. Personalized Content Retrieval in Context Using Ontological Knowledge. IEEE Transactions on Circuits and Systems for Video Technology 17(3), March 2007, pp. 336-346.

[29] A. Vetro, C. Christopoulos, T. Ebrahimi (Eds.). Special Issue on Universal Multimedia Access. IEEE Signal Processing Magazine 20(2), March 2003.

[30] A. Vetro. Transcoding, Scalable Coding and Standardized Metadata. 8th International Workshop on Visual Content Processing and Representation (VLBV 2003). Madrid, Spain, September 2003. Springer Verlag LNCS Vol. 2849, 2003, pp.15-16.

[31] A. Vetro and C. Timmerer. Digital Item Adaptation: Overciew of standardization and research activities. Special issue on MPEG-21, IEEE Transactions on Multimedia 7(3), June 2005, pp. 418-426.

[32] F. López and J. M. Martinez. Multimedia Content Adaptation Modelled as a Constraints Matching Problem with Optimisation. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2007). Greece, 2007, pp. 82-85.