A Collaborative Filtering Approach to Personalized Interactive Entertainment using MPEG-21

Phivos Mylonas, Giorgos Andreou and Kostas Karpouzis
National Technical University of Athens
School of Electrical and Computer Engineering
Department of Computer Science
Image, Video and Multimedia Laboratory,
Zographou Campus, Athens, Greece, 157 73

Abstract. In this chapter we present an integrated framework for personalized access to interactive entertainment content, using characteristics from the emerging MPEG-21 standard. Our research efforts focus on multimedia content presented within the framework set by today's movie content broadcasting over a variety of networks and terminals, i.e. analogue and digital television broadcasts, video on mobile devices, personal digital assistants and more. This work contributes to the bridging of the gap between the content and the user, providing end-users with a wide range of real-time interactive services, ranging from plain personalized statistics and optional enhanced in-play visual enhancements to a fully user- and content-adaptive platform. The proposed approach implements and extends in a novel way a well-known collaborative filtering approach; it applies a hierarchical clustering algorithm on the data towards the scope of group modelling implementation. It illustrates also the benefits from the MPEG-21 components utilization in the process and analyzes the importance of the Digital Item concept, containing both the (binary) multimedia content, as well as a structured representation of the different entities that handle the item, together with the set of possible actions on the item. Finally, a use case scenario is presented to illustrate the entire procedure. The core of this work is the novel group modelling approach, on top of the hybrid collaborative filtering algorithm, employing principles of taxonomic knowledge representation and hierarchical clustering theory. The outcome of this framework design is the fact that end-users are presented with personalized forms of multimedia content, thus enhancing their viewing experience and creating more revenue opportunities to content providers.

Keywords. Personalization, Collaborative Filtering, MPEG-21, Digital Item, Network Management Adaptation

Introduction

In the new era of interactive public and home entertainment, a new generation of content consumers has been born and is currently confronted with a series of technological developments and improvements in the digital multimedia content realm. Their expectations are high and the need for a high quality service problem handling is more stressful than ever. At the same time, digital video is the most demanding and complex data structure, due to its large amounts of spatiotemporal interrelations;

efficient manipulation of visual media is currently not considered a trivial task. Multimedia standards such as MPEG-4 [1], [3], [4] and MPEG-7 [2], provide important functionalities for coding, manipulation and description of objects and associated metadata; however, personalized filtering of the content, provided it is accompanied by corresponding metadata, is out of the scope of these standards, motivating heavy research efforts and the emerge of a new standard, i.e. MPEG-21 [5], [6].

Domains characterized by inherent dynamics, such as movie collections and broadcasting, make the above expectations even higher; thus, broadcasting corporations and organizations need to preserve or build up their competitive advantage, seeking new ways of creating and presenting enhanced content to their new, demanding content consumers. From the emerging mobile devices point of view, third-generation (3G) services provide the ability to transfer simultaneously both voice data (i.e. a telephone call) and non-voice data, such as downloading of a movie. In marketing 3G services, video telephony has often been used as the killer application for 3G. Both aspects will greatly enhance the multimedia content transmission and consuming potential of mobile devices. Consequently, the need for innovative services over 3G networks is large, in order to facilitate wide take up of the new technology by their end-users.

Multimedia content retrieval and filtering in the last decade has been influenced by the important progress in numerous fields such as digital content production, archiving, multimedia signal processing and analysis, as well as information retrieval. One major obstacle, though, such systems still need to overcome in order to gain widespread acceptance, is the semantic gap [7]. This refers to the extraction of the semantics of multimedia content, the interpretation of user information needs and requests, as well as to the matching between the two. This obstacle becomes even harder when attempting to access vast amounts of multimedia information and metadata contained within a movie.

In current research activities it is becoming apparent that offering of integrated personalized interactive services upon diverse and possibly heterogeneous – pre-existing – multimedia content will only be feasible through novel techniques and methodologies. In [30] for instance, a personalized content preparation and delivery framework for universal multimedia access is introduced. On the other hand, [31] focuses on a novel approach to support adaptive services for multimedia delivery in heterogeneous wireless networks.

Moreover, motion pictures (movies) continue to attract interest and are among the most popular media attractions in the world today. Consequently, multimedia applications developed for them have a huge potential market impact. Among these applications are the provision of enhanced content, statistics, dynamic interactive content and optionally advertisements. In this framework, our work targets the provision of enhanced content and statistics to provide a friendly, easily assimilated interface, as well as dynamic, personalized interactive content and advertisements to enable the user to interact with the content, thus enhancing the user experience by providing further information on the specific movie.

Our efforts resulted in an integrated framework, offering transparent, personalized access to heterogeneous multimedia content, using characteristics from the emerging MPEG-21 standard. Although recently applied in the sports domain [38], focusing on a network, device and user independent solution, this approach contributes towards bridging the gap between the semantic nature of user needs and raw multimedia documents - as expressed by movies, serving as a management mediator between end-users and movie repositories. Its core contribution relies on the fact that it provides a personalized delivery of content over heterogeneous networks and terminals, using the core functionality of the MPEG-21 standard and providing the missing link for an integrated personalized interactive experience. The latter is achieved by utilizing the notion of an MPEG-21 Digital Item [39], using it to encapsulate personalization-useful information at the multimedia content level and not at the level of terminal or system.

In this context, a user is any entity that interacts with or makes use of a Digital Item. A hybrid collaborative filtering method is then applied, based on this unified knowledge model and multimedia documents (i.e. movies) are clustered according to their ratings through clustering on their features. Future user requests are then analyzed and processed to retrieve movies from the framework's repository, according to the underlying user preferences. This chapter presents an integrated approach in the framework of the MPEG-21 standard to establish the necessary infrastructure to support the virtual value chain for personalized interactive entertainment events broadcasting over wireless, cable and digital networks, offering valuable and revenue-building services.

It should have been obvious by now that watching multimedia entertainment content at home or in public tends clearly to be a social activity. So, adaptive content providers and consumers need to adapt content to groups of users rather than to individual users. In this chapter, we discuss a hybrid strategy for combining individual user models to adapt to groups, which is basically inspired by the Social Choice Theory [37], i.e. how humans select a sequence of items (e.g. movies) for a group to watch, based on data about the individuals' preferences. The latter offers the possibility of personalized viewing experiences, based on features that pre-exist in the information accompanying each multimedia item/movie. In our framework, information on movie characteristics is derived from the Internet Movie Database (IMDB) [35].

The IMDB consists of the largest known single accumulation of data on a vast amount of multimedia content, including individual films (together with their complete cast and crew listings), television programs (including complete cast and crew listings), direct-to-video product and videogames reaching back to their respective beginnings, and worldwide in scope. Wherever possible, the information goes beyond simple screen or press credits to include uncredited personnel involved, either artistically or technically, in the production and distribution, thus aiming at completeness of detail. Furthermore, a collateral database of all persons identified in the product database exists, including biographical details and information, such as theatrical appearances, commercial advertising appearances, etc. Information is largely provided by a cadre of volunteer contributors and is considered to be the most accurate and up-to-date multimedia content database at the time of the writing of this book chapter.

Adapting this kind of multimedia content to individual viewers is a topic in itself, and a lot of research has already been done. Moreover, different domains have been identified in which a personalization process would have a great impact, such as education [32], advertising [33], and electronic program guides [34]. This research tends to build on decades of work on content-based and social filtering. As already discussed, herein we focus our efforts on exploring an even more difficult issue: adaptation of multimedia content to a group of viewers. We believe this to be essential for interactive multimedia content viewing as, in contrast to the plain use of personal computers or televisions, multimedia content viewing is largely a family or social activity. In this context, recommender systems are a special class of personalized systems that aim at predicting a user's interest on available products and services by relying on previously rated items or item features. Human factors associated with a user's personality or lifestyle, although potential determinants of user behaviour are rarely considered in the personalization process. It is a fat, that, the concept of lifestyle can be incorporated in the recommendation process to improve the prediction accuracy by efficiently managing the problem of limited data availability.

The structure of this chapter is as follows: Section 1 provides a high level overview of the proposed framework, focusing on its structure and data models. It also describes the notions of Collaborative Filtering (CF) and Hierarchical Clustering, together with a brief introduction to the MPEG-21 standard. Subsection 1.5 provides a detailed use case scenario, defining the scope of this work. Section 2 describes in detail the proposed hybrid collaborative filtering approach, based on hierarchical clustering applied on the movies' features. Continuing, section 3 discusses the basics of the MPEG-21 Digital Item utilization, followed by the corresponding resource adaptation within the proposed framework. In section 4 conclusions are drawn and some future work aspects of this work are also discussed.

1. Overview of the proposed framework

1.1. Framework architecture

The proposed framework is illustrated in Figure 1 and involves a variety of user terminals and networks, such as Personal Digital Assistants (PDA) and Personal Computers (PCs) over TCP/IP networks, Set-Top-Boxes (STB) over Local Cable TV networks, High Definition Digital Television Sets (HDTV) and networks, as well as Mobile Devices over UMTS, GPRS, or GSM networks. Given the diversity and singularity that characterize each type's multimedia content receiving, processing and displaying capabilities, specific care must be taken for its adaptability and presentation to the end-user, as depicted by the *Content Preparation and Adaptation* component, as well as intelligent content customization (e.g. subtitling and/or dubbing), as depicted by the *Information Merging Unit*. Prior *Visual Enhancements Engine* includes preparation of (optional) advertising content, preparation of non-standard events to be offered during a broadcast (e.g. real-time movie trivia) and designing descriptive templates for the display of optional enhanced content at transmission time. Content adaptation to any kind of end-user terminal is performed according to the so called "create once publish everywhere" principle, adapted to the targeted network and terminal prior to transmission, to allow for efficient display and manipulation on the end-user side.

On top of that, reusability of the content is considered a prerequisite in modern content management applications, although still an open and difficult issue to tackle; creating reusable multimedia content demands well-structured and efficient data models, together with highly-refined content repositories on a massive scale and is an expensive process that requires careful planning and design. Besides this, since the integrated framework handles, encodes and presents multimedia content coming from different vendors, the respective intellectual property rights (IPR) must also be retained throughout the complete process. These two additional requirements can be dealt with successfully via the inclusion of concepts presented within the emerging MPEG 21 framework, as discussed in the following sections of this chapter. Figure 1 shows only an overview of the higher-level information flow, between the different framework components. The components that undertake the task of collecting, packaging and delivering of multimedia data have been collectively enhanced to cater for the aforementioned provisions in the final content, which is transmitted to the end-user via the *Video Content Transmission* component. The reader is encouraged to find a detailed description of the complete process level architecture framework in [8]

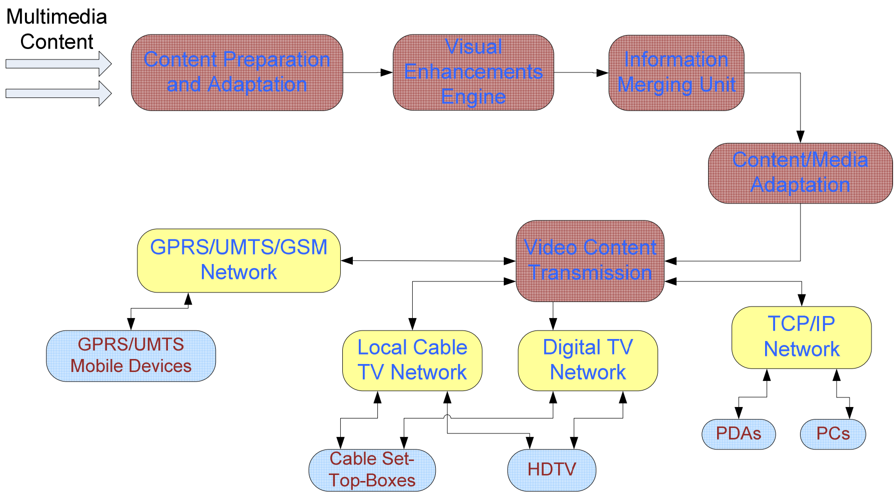

Figure 1. Overview of the proposed framework.

The basic idea is that content is adapted to the different terminals and transmission networks targeted by the proposed framework and then delivered via the respective transmission channels. In the case of TCP/IP and GPRS/UMTs/GSM broadcast, the video is streamed in MPEG-4 over an MPEG-2 Transport Stream. The video resolution is then reduced to fit the lower transmission and playback capabilities of mobile terminals. Since the targeted receiver architectures offer different degrees of media delivery, interactivity and responsiveness, it is essential to break down both the captured and synthesized material to match the relevant device. As a result, different versions of the content are prepared for delivery. In the next subsections we briefly present our proposed methodology guidelines, in order to enable personalization aspects, based on the content preparation principles discussed previously.

1.2. Collaborative Filtering

In the context of bridging the gap between the content and the user and providing personalized interactive services, we implement and extend a widely-known Collaborative Filtering (CF) technique. Collaborative Filtering is the method of making automatic predictions or filtering about the interests of a user by collecting preference information from a larger pool of users [21]. The underlying assumption in all CF approaches is that users who agreed in the past, tend to agree again in the future. In the case of a collaborative filtering system for multimedia content preferences one could make predictions about which movie a user should like given a partial list of that user's preferences. These predictions are specific to the user, but use information gleaned from many users. This differs from the more simplistic approach of giving an average score for each movie of interest, for example based on its number of favouring votes.

Many variations of collaborative filtering algorithms and systems exist; however, most of them usually take two steps:

1. Look for users who share the same rating patterns with the active user (the user who the prediction is for).
2. Use the ratings from those like-minded users found in step 1 to calculate a prediction for the active user.

Alternatively, item-based collaborative filtering [16], [17] popularized by Amazon.com (i.e. “users who bought x also bought y”) proceeds in an item-centric manner:

1. Build an item-item matrix determining relationships between pairs of items
2. Using the matrix, and the data on the current user, infer his/her preference

In the age of information explosion such techniques can prove very useful as the number of items in multimedia content (such as music, movies, news, web pages, etc.) have become so large that a single person cannot possibly view them all in order to select relevant ones. On the other hand, relying solely on a scoring or rating system which is averaged across all users ignores specific demands of a user, and its outcome may be particularly poor in tasks where there is large variation in interest, like movies or music recommendation. Consequently, other methods to combat information explosion must aid in the process and in the scope of this work we focused on one of them, i.e. hierarchical data clustering.

1.3. Hierarchical Clustering

The essence of clustering data is the classification of similar objects into different homogeneous groups, based on the values of their attributes. More precisely, data clustering is the partitioning of a given data set into subsets (clusters), so that the data in each subset share some common trait according to some defined distance measure. It is a problem that is related to various scientific and applied fields and has been used in science and in the field of data mining for a long time, with applications of techniques ranging from artificial intelligence, machine learning, data mining and pattern recognition, to image analysis, bioinformatics, databases and statistics [18].

There are different types of clustering algorithms for different types of applications and a common distinction is between *hierarchical* and *partitioning* clustering algorithms. Hierarchical algorithms find successive clusters using previously established clusters, whereas partitioning algorithms determine all clusters at once. Although hierarchical clustering methods are more flexible than their partitioning counterparts, in that they do not need the number of clusters as an input, they are less robust in initial error propagation and computational complexity issues and thus must be used with caution and under specific circumstances, as depicted in the following. In general, clustering of data is still considered an open issue, basically because it is difficult to handle in the cases that data is characterized by numerous measurable features, as in the case of movie features.

1.4. MPEG-21 Digital Items

The basic architectural concept in MPEG-21 is the Digital Item. Digital Items are structured digital objects, including a standard representation, identification and metadata. They are the basic unit of transaction and distribution in the MPEG-21 framework [22]. More concretely, a Digital Item is a combination of resources (such as videos, audio tracks, images, etc), metadata (such as descriptors, identifiers, etc), and structure (describing the relationships between resources). The second part of MPEG-21 (ISO/IEC 21000-2:2003) specifies a uniform and flexible abstraction and interoperable schema for declaring the structure and makeup of Digital Items. Digital Items are declared using the Digital Item Declaration Language (DIDL) and declaring a Digital Item involves specifying its resources, metadata and their interrelationships. In this context, complex digital objects, as the ones containing multimedia content feature information used in the presented hybrid CF approach, may be declared using the notion and language of a Digital Item. DIDL language defines the relevant data model via a set of abstract concepts, which form the basis for an XML [23] schema that provides broad flexibility and extensibility for the actual representation of compliant data streams.

The usage of the MPEG-21 Digital Item Declaration Language to represent such complex digital objects, has introduced benefits to the proposed framework in two major areas: The management of the initial content presentation and the management and distribution of multimedia content, such as video, images and metadata. The platform allows the creation of predefined templates during the planning process before broadcasting; these templates form the Initial Scene that is used to generate the initial MPEG-4 scene. During the broadcasting phase, templates are used in order to control the real time updates of the MPEG-4 content. Having all the information packaged in one entity, i.e. initial scene, customization points, etc. brings the benefit of reduced complexity data management.

Furthermore, the benefit from the adoption of MPEG-21 is that every Digital Item can contain a specific version of the content for each supported platform. The dynamic association between entities reduces any ambiguity over the target platform and the content. Having all the necessary information packaged in one entity enables the compilation and subsequent adaptation of a Digital Item to be performed only once (during its creation) and not on a per-usage basis, thereby effectively eliminating the need for storage redundancy and bringing the benefit of reduced management and performance complexity in the Information Repository. The adopted MPEG-21 concepts and their structure are described in detail in subsequent sections of this chapter.

1.5. A Use Case Scenario

At this point let us assume that the multimedia content offered to the end-users of the proposed standalone system contains a set of movies to choose from. These can be movies whose main genre is comedy, drama, science fiction, etc. For the sake of simplicity, we utilize only a part of the IMDB movie information. More specifically, we take into consideration only the subset of 14 movie attributes presented in 2:

#	Feature
1.	Actor
2.	Actress
3.	Director
4.	Genre
5.	Language
6.	Location
7.	MPAA Rating
8.	Plot summary
9.	Producer
10.	Rating
11.	Release Date
12.	Running Time
13.	Title
14.	Writer

Table 1. IMDB movie features selection

Let the end-users have preference ratings over the set of movies, either for a specific movie or for a group of movies (i.e. cluster of movies), in the following 1-10 scale-based manner: 1 is used to denote a really negative preference, i.e. “really hate”, whereas 10 denotes a “really like” preference. The basic problem is which movies the content provider should offer to new users, given the preferences of existing users over the set of offered movies.

A simplified example of this situation may be given as follows: three end-users, John, George and Mary are already watching their favourite movies: John has invited his friends at home to watch a comedy film in his new home cinema, George is travelling by train and watches his favourite drama movie in his PDA and Mary is waiting for her turn in the doctor’s office watching a drama movie on a Set-Top-Box. All three of them have established their user preferences for a set of ten (randomly selected) movies (A to J), that include the three that are currently viewing. As expected, each end-user has a different view on the quality of the 10 selected movies and rates them according to his/her subjective criteria. A new user, Tom, opens his personal computer and requests from the content provider the top movies, according to system’s user ratings, to select from.

2. Personalization and Filtering: a Hybrid Approach

One of the technical novelties introduced in the proposed framework is the handling of its users in a personalized manner, by building different profiles according to their preferences. The system is able to provide each user personalized multimedia content according to his/her specific user profile; a functionality provided considering a hybrid collaborative filtering methodology, based on hierarchical clustering on content information acquired by all participating content material. Current section of this work presents the design and implementation of the profile-based framework, which matches content to the context of interaction, so that the latter can be adapted to the user's needs and capabilities.

In this context, we introduce a novel hybrid collaborative filtering approach, based on multimedia content clustering. More specifically, we apply traditional data mining techniques, such hierarchical clustering on the multimedia content itself (i.e. movies), according to a predefined set of features. This set includes movies' characteristics, such as movie genre, filming date, movie type, etc and is distinctive of the content. All this information is encapsulated within the Digital Item concept of MPEG-21, to ensure interoperability and robustness of the overall approach, as well as network and terminal independency. The latter is achieved through the adoption of the MPEG-21 standard and the lack of a single centralized system database; quite on the contrary, all necessary information is content- and user-centric, decentralized to all participating user terminals.

In the remaining of this section, we present a brief overview of our hybrid collaborative filtering clustering technique, together with a detailed description of the proposed algorithm.

2.1. Hierarchical Clustering Algorithm

In order for the proposed framework to provide the discussed kind of personalized access to interactive entertainment, a number of steps need to take place. The first step in identifying the suitable set of top ranked movies in the system is to cluster them according to the set of features under consideration. This step is necessary in order to identify homogeneous patterns in the movie data set that will aid in the personalization process in terms of selection speed and quality. As already discussed, the main problem a clustering technique is asked to solve is the identification of homogeneous groups of objects based on the values of their attributes. In general, this is a difficult problem to tackle, related to various scientific fields, especially when clustering is applied to user modelling [9], [10]. The problem gets more and more challenging, as input space dimensions become larger and feature scales are different from each other, as is the case of our framework. In particular, a consideration of the original set of movie characteristics, as described in the Internet Movie Database [20] and Movielens [19] as input space, results into a large number of unique features to be taken into consideration when performing clustering on this kind of multimedia content, i.e. movies.

The best way to go in this direction is to use a hierarchical clustering algorithm, which is able to tackle such a large scale of features [11], [12]. Although such a method does not demand the number of clusters as input, still it does not provide a satisfactory framework for extracting meaningful results. This is mainly due to the "*curse of* *dimensionality*" that dominates such an approach, as well as the inevitable initial error propagation and complexity along with data set size issues. The "*curse of dimensionality*" is a term applied to the problem caused by the rapid increase in volume associated with adding extra dimensions to a mathematical space. As an example consider that having 100 observations covering the one-dimensional unit interval [0,1] on the real line provides a quite well performance; however considering the corresponding 10-dimensional unit hypersquare, the 100 observations would be isolated points in a huge empty space. To get similar coverage to the one-dimensional space would now require 10^{20} observations, which is at least a massive undertaking.

In this context and in order to increase the robustness and reliability of the whole clustering step of our system, the use of an unsupervised extension to hierarchical clustering in the means of feature selection is evident [11]. The results of the application of this clustering to only a portion of the movie dataset in question are then refined and extended to the whole dataset. In this approach we follow the standard hierarchical clustering algorithm structure:

1. We start from turning each movie into a singleton m_i (i.e. a cluster containing a single movie).
2. Then, for each pair of clusters m_1, m_2 we calculate their distance $d(m_1, m_2)$.
3. Identifying the smallest distance amongst all possible pairs of clusters results into the merging of this pair.
4. The above described process is repeated from step 2, until a meaningful clustering termination criterion is satisfied; the termination criterion most commonly used is a meaningful threshold on the clusters' distance value $d(m_1, m_2)$.

As it is typical in cases where the input space dimensions are large, the Euclidean distance is considered to be the best distance measure used [13]. Still, this is not always the case, due to the nature of the individual features; consequently a selection of meaningful features needs to be performed, prior to calculating the distance $d(m_1, m_2)$ [14]. On the one hand, one feature might be more important than others, while at the same time all features are useful to some degree. As a result and according to these principles, an additional meaningful weighting of features is followed within our approach. The key element of the above algorithm is the ability to define a unique distance among any pair of clusters, given the input space and the clustering features. More formally, letting m_1 and m_2 be two clusters of movies, we propose the following distance measure when considering just the *i*-th feature:

$$f_i(m_1, m_2) = \sum_{i \in \mathbb{N}_F} \sqrt[\mu]{\frac{\sum_{x \in m_1, y \in m_2} r_i(x_i, y_i)^{\mu}}{|m_1||m_2|}} \quad (1)$$

where $r_i, i \in \mathbb{N}_F$ is the metric that compares the i-th feature, F the overall count of features, $|m_1|$ the cardinality of cluster m_1 and μ a constant. Obviously, $\mu = 1$ approaches the mean value and $\mu = 2$ yields the Euclidean distance. The overall distance between m_1 and m_2 is calculated as:

$$d(m_1, m_2) = \sum_{i \in N_F} x_i(m_1, m_2)^{\lambda} f_i(m_1, m_2) \qquad (2)$$

where x_i is the degree to which i, and therefore f_i, is included in the soft selection of features, $i \in N_F$ and λ is a constant. When $\lambda = 1$ the solution is trivial and the feature that produces the smallest distance is the only one selected with degree equal to 1. For the sake of simplicity, the interested reader is encouraged to read the detailed approach on the issue that can be found in [15]. Finally, the above described clustering approach creates crisp clusters of movies and does not allow for overlapping among the detected clusters. Thus, it forms the basic procedure, with the aid of which movies are automatically categorized to a distinct group class that will be used during the collaborative filtering step of this approach. The movies' clustering characterizes the users' behaviour and future interests and choices of multimedia content.

2.1.1. Hierarchical Clustering Algorithm Implementation

In this section, we examine the implementation of the proposed hierarchical clustering algorithm using system's movie data set and the Euclidean distance measure. The clustering algorithm has been applied to a small portion of the data set, namely a 10% of the overall movies; it contained 100 elements (movies), characterized by 14 meaningful features. These features have been considered appropriate for the personalization process and were selected a priori by a group of experts. Identified clusters define specific interests and preference information. These clusters are useful in producing collaborative recommendations of the multimedia content to the end-users at the later request stage, as described in section 3. Results are shown in Table 2, Table 3 and Table 4, whereas the letters inside parenthesis separated by punctuation marks denote the movies belonging to its cluster in each step.

Performing the initial clustering on a mere 10% subset is not only more efficient computationally wise, it is also better in the means of quality and performance, when compared to the approach of applying the hierarchical process to the whole data set. Although clustering over this 10% of the data set resulted in different possible identifiable clusters, optimal results have been obtained for a number of nine clusters, as indicated in the following tables, where clustering results are presented for three variations of output clusters (3, 5 and 9):

More specifically, Table 2 presents the clustering results of 100 learners. The hierarchical clustering algorithm terminated by the time it reached a threshold of 3 clusters. The first cluster comprises of 17 learners, namely 2 belonging to the *Experts* class, 6 to the *Beginners* class and 9 to the *Advanced* class. The corresponding percentage distribution clearly indicates that the 9 *Advanced* dominate the first cluster. The second cluster consists of 38 learners: 11 *Experts*, 2 *Beginners* and 25 *Advanced*. *Advanced* are dominant in this cluster as well, whereas in terms of percentages their domination is confident (i.e. 65,79%). Finally, the third cluster contains 45 learners, 14 of whom are *Experts*, 1 is *Beginner* and 30 are *Advanced*. Domination of *Advanced* is more indicative, since a 66,67% gives them a clear advantage.

Clusters	Elements
1st	17
2nd	38
3rd	45

Table 2. 100 movies clustering results – 3 clusters

Table 3 presents the clustering results of the same 100 movies; however a new threshold of 5 clusters terminates the clustering algorithm earlier. In this case, all clusters contain lesser learners, in comparison to the prior case. The first cluster consists of 11 movie items. The corresponding percentage distribution clearly indicates that the 7 movies dominate the first cluster. The same applies to the second cluster with 14 movies as well. Third cluster contains 19 movies, fourth cluster contains 25 movies and fifth cluster contains 31 movies.

Clusters	Elements
1st	11
2nd	14
3rd	19
4th	25
5th	31

Table 3. 100 movies clustering results – 5 clusters

Continuing, in Table 4 are presented the results of the clustering step terminating in 9 movie clusters:

Clusters	Elements
1st	6
2nd	7
3rd	11
4th	12
5th	11
6th	17
7th	9
8th	19
9th	8

Table 4. 100 movies clustering results – 9 clusters

2.2. Collaborative Filtering

Our CF algorithm recommends movies to the active user based on the ratings to the previously clustered movie titles of *n* other users. It it summarized in the following principles:

i) Let the set of all movie titles be *M* and the rating of user *i* for title *j* as $r_i(j)$. The function $r_i(j): M \rightarrow \Re \cup \{\perp\}$ maps titles to real numbers or to $\perp$, the symbol for "no rating."
ii) Denote the vector of all of user *i*'s ratings for all titles as $r_i(M)$
iii) Denote the vector of all of the active user's ratings as $r_a(M)$.
iv) Define $NR \subset M$ to be the subset of titles that the active user has not rated, and thus for which we would like to provide predictions. That is, title *j* is in the set *NR* if and only if $r_a(j) = \perp$.
v) Then the subset of titles that the active user *has* rated is *M-NR*.
vi) Define the vector $r_i(S)$ to be all of user *i*'s ratings for any subset of titles $S \subseteq M$, and $r_a(S)$ analogously.
vii) Finally, denote the matrix of all users' ratings for all titles simply as *r*. In general terms, a collaborative filter is a function *f* that takes as input all ratings for all users, and outputs the predicted ratings for the active user:

$$r_a(NR) = f(r_1(M),\ r_2(M),\ \ldots,\ r_n(M)) = f(r) \qquad (3)$$

where the $r_i(M)$'s include the ratings of the active user.

2.2.1. Collaborative Filtering Implementation

End-users have preference ratings over the set of clustered movies in the following 1-10 scale-based manner: 1 is used to denote a really negative preference, i.e. "really hate", whereas 10 denotes a "really like" preference. The basic problem is which movies should the content provider offer to new users, based on the ratings of existing users. Following this principle, we provide an example of 3 individual user ratings over the identified 9 clusters on the subset of 100 movies, as depicted in Table 5:

	MC1	MC2	MC3	MC4	MC5	MC6	MC7	MC8	MC9
User 1	10	4	3	6	10	9	6	8	8
User 2	1	9	8	9	7	9	6	9	3
User 3	10	5	2	7	9	8	5	6	7

Table 5. Example ratings for a group of viewers – MC: Movie Cluster

Many strategies, also called "social choice rules" or "group decision rules" have been devised for reaching group decisions given individual opinions. The one followed herein originates from the Social Choice Theory and will be illustrated with the example introduced above. Table 6 shows the "group preference ranking/rating" resulting from the strategy, a sequence indicating in which order movie clusters would be chosen, when a new end-user requests a movie rating. In this approach, utility values for each alternative are used, instead of just using ranking information as in other approaches (e.g. in the “plurality voting” approach). More specifically, ratings are added, and the larger the sum the earlier the alternative appears in the final movie rating sequence. This strategy is widely spread and used also in a variety of systems and approaches, such as multi-agent systems [36].

	MC1	MC2	MC3	MC4	MC5	MC6	MC7	MC8	MC9
User 1	10	4	3	6	10	9	6	8	10
User 2	1	9	8	9	7	9	6	9	3
User 3	10	5	2	7	9	8	5	6	7
Group	21	18	13	22	26	26	17	23	20
Group Rating	(MC5, MC6), MC8, MC4, MC1, MC9, MC2, MC7, MC3								

Table 6. Example group ratings for a new user – MC: Movie Cluster

2.3. Personalization using MPEG-21 concepts

According to the previously analyzed methodology, the system provides end-users with the possibility to see only movies and information about movies that they are interested in. One flexible way to perform content personalization is to filter the content that is streamed to the client. In the case of a STB display, since the same content is broadcast to all clients, filtering should occur at the client side, i.e. on the STB Mary is watching a drama movie before its presentation. The MPEG-21 framework is used for personalization and content filtering in the following way: Mary’s STB contains an MPEG-21 DIA Description that specifies her user preferences on content. When her user terminal receives multimedia content, this is filtered according to its genre and Mary’s user preferences indicated in the DIA Description. The main issue is to find a way to transport synchronously multimedia content and its associated metadata indicating its genre, in order to make sure that the multimedia content is not received before its description. One way to achieve this is by grouping the multimedia content and its genre within a DID, and to stream the complete DID to the clients. In the case of Mary, it safe to assume that this DID indicates that the multimedia content belongs to the genre “Drama”. Obviously, according to the user preferences of a “Comedy”-based DIA Description, the multimedia content will be filtered out by the client terminal and therefore not displayed, whereas in the case of a “Drama”-based DIA Description preferences, it will be promoted and presented to Mary. The same applies to John and George at home and on a train, respectively.

3. MPEG-21 Digital Item Utilization

3.1. Digital Item Declaration

The task of creating a robust architecture framework for creating and delivering of diverse multimedia content has been in the past and currently continues to be an ambitious mission. MPEG-21 introduced the Digital Item (DI), a new interoperable unit for multimedia delivery and transaction. As in any environment that proposes to facilitate a wide range of actions involving "Digital Items", there is a need for a very precise description for defining exactly what constitutes such an "item". The basic concept of a DI is essentially a container for all kinds of metadata and content and at the first glance might look partially similar to work undertaken in other fields, such as e-learning [24]. However, in MPEG-21, a complete, flexible and rich delivery framework based around a more versatile DI specification has been standardized at a higher level framework.

The general structure of a DI is provided by a Digital Item Declaration (DID) [25]. A DID is a document that specifies the makeup, structure and organization of a DI. The DID formally expresses and identifies the content and the metadata, that are considered to be the components of the DI. Further, the DID binds together groups of resources and metadata, i.e. movies together with their characteristic features. In the case of the proposed framework, the DIs follow the standardized MPEG-21 principle elements, where items like genre and/or user ratings are grouped together into components that are grouped into a container; an example of a DI declaration code is depicted in the following Figure 2:

```
<?XML VERSION="1.0" ENCODING="utf-8"?>
<DIDL xmlns="urn:mpeg:mpeg21:2002:01-DIDL-NS"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="urn:mpeg:mpeg21:2002:01-DIDL-NS didl.xsd"
xmlns:dii="urn:mpeg:mpeg21:2002:01-DII-NS">
<ITEM ID="Movie #1">
  <DESCRIPTOR>
    <STATEMENT TYPE="text/plain">Title: The Devil Wears Prada</STATEMENT>
  </DESCRIPTOR>
  <ITEM>
    <DESCRIPTOR>
      <STATEMENT TYPE="text/plain">
        Date: 2006
      </STATEMENT>
    </DESCRIPTOR>
  </ITEM>
  <ITEM>
    <DESCRIPTOR>
      <STATEMENT TYPE="text/plain">
         Plot Outline: A naive young woman comes to New York and scores a job as the
         assistant to one of the city's biggest magazine editors, the ruthless and cynical
         Miranda Priestly.
      </STATEMENT>
    </DESCRIPTOR>
  </ITEM>
  <ITEM>
    <DESCRIPTOR>
      <STATEMENT TYPE="text/plain">
        User rating: 7.0/10.0
      </STATEMENT>
    </DESCRIPTOR>
  </ITEM>
</ITEM>
</DIDL>
```

Figure 2. Example of a Digital Item Description.

The DID opens with the XML namespace declarations familiar to users of XML and the root DIDL element. Within this DIDL element we have a single Item which has an id attribute, allowing external or internal referencing of the Item. Item identifiers of any level (i.e. DII Identifier elements) are not included here for brevity. Human readable text Descriptor/Statement combinations (such as the title, date, genre, plot outline and user rating of the movie) follow in order to provide interoperability.

3.2. Digital Item Identification

Up to now, digital identification in multimedia documents came in usual ways, such us nested digital sign in file headers and declared types in file extensions. The lack of homogeneity as it was developed by the diversity of media formats and the richness of multimedia content itself provoked as much confusion as opacity in the multimedia chain. On the contrary, the MPEG-21 vision of interoperability and transparency, leads the way to fill the gap between the different technologies. As already discussed, MPEG-21 objects, like “item” and “user,” become the main entities, in order to describe the general idea in a more transparent and less complicated manner.

The role of Digital Item Identification (DII) is not only to propose the way to identify DIs in a unique manner, but also to distinguish different types of them. These Identifiers are placed in a specific part of the DID, which is the statement element, and they are associated with DIs: DIs are identified by encapsulating uniform resource identifiers, which are a compact string of characters for identifying an abstract or physical resource. The elements of a DID can have zero, one or more descriptors; each descriptor may contain a statement which can contain an identifier relating to the parent element of the statement. Besides the references to the resources, a DID can include information about the item or its parts. An example about the metadata that a movie could have in MPEG-21 within our framework is visualized in

Figure 3. As obvious from the figure, it is necessary for DII to allow differentiating between the different schemas that users can use to describe content in general. Consequently, MPEG-21 DII uses the XML [23] mechanism to achieve this.

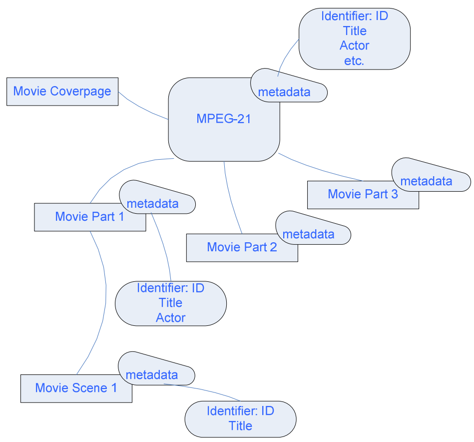

Figure 3. Visualization example of a movie Digital Item implementation.

3.3. Multimedia resource adaptation

The focus of resource adaptation is the framework of DIA (Digital Item Adaptation), where messages between servers and end-users are in the form of XML documents with URL links to resources or encoded binary data. In the case of linked resources, a Digital Resource Provider decides which variation of the resource is best suited for the particular user, based on the user’s terminal capabilities, the environment in which the user is operating and the available resource variations. In our use case scenario, for example, where George views his favourite drama movie travelling on a train, i.e. a streaming media resource, adaptation will depend on the available bandwidth, screen size, audio capabilities and available viewer software in his PDA terminal, all part of an automated process, as capabilities and preferences should be automatically extracted and enforced.

DIA is the key element in order to achieve transparent access to distributed advanced multimedia content, by shielding end-users like George from network and terminal installation, management and implementation issues. The latter enables the provision of network and terminal resources on demand to form user communities where multimedia content can be created and shared, always with the agreed/contracted quality, reliability and flexibility, allowing the multimedia applications to connect diverse sets of users, such that the quality of the user experience will be guaranteed.

Towards this goal the adaptation of DIs is required. In the described platform dynamic media resource adaptation and network capability negotiation is especially important for the mobile paradigm (the George/PDA paradigm) where users (George) can change their environment (i.e. locations, devices etc) dynamically (e.g. get off the speeding train or request the same content for his mobile phone as well). MPEG-21 addresses the specific requirements by providing the DIA framework, for scalable video streaming [25]. The DIA framework, specifies the necessary mechanisms related to the usage environment and to media resource adaptability. This approach was the one adopted for the proposed platform. Alternative approaches to this issue may be the HTTP and RTSP based [27], [26], [29] or the agent based content negotiation mechanisms [28].

The DIA framework, regarding resource adaptation, includes the Usage Environment Description Tools (i.e. User Characteristics, Terminal Capabilities, Network Characteristics, Natural Environment Characteristics) and the Digital Item Resource Adaptation Tools (i.e. Bitstream Syntax Description - BSD, Adaptation QoS, Metadata Adaptability). The latter are the main tools, which enable resources adaptation. BSD language provides information about the high level structure of bitstreams so that streaming can be modified accordingly to this information. Adaptation QoS schema provides the relationship between QoS parameters (e.g., the current network interface bandwidth in the case of Mary or George’s PDA computational power) and the necessary adaptation operations needed to be executed for satisfying these parameters. The associated video or media quality, which is the outcome of the adaptation procedure, is also included as parameter in the adaptation schema.

4. Conclusions

The core contribution of this work has been the provision of an integrated framework for personalized access to heterogeneous interactive entertainment multimedia content, using characteristics from the emerging MPEG-21 standard. It contributed to the bridging of the gap between the raw content and the end-user over a variety of networks and terminals. This is accomplished by implementing a novel collaborative filtering approach. It utilized a hierarchical clustering algorithm towards the scope of group modeling implementation, illustrating at the same time the benefits from the use of MPEG-21 standard components, such as the Digital Items. Finally, a real-life use case scenario is presented to illustrate the entire procedure.

The methodology presented in this book chapter can be exploited towards the development of more intelligent, efficient and personalized multimedia content access systems, thus enhancing the viewing experience of end-users. In order to verify its efficiency, we plan to have it thoroughly tested in the framework of a future multimedia retrieval, personalization and filtering Digital TV application. Another interesting perspective for future work is the utilization of end-users' usage history. Additionally, design and efficiency issues improvement of knowledge representation and hierarchical clustering process presented herein lies within our first research priorities.

References

[1] Battista S., Casalino F., Lande C., "MPEG-4: A Multimedia Standard for the Third Millenium, Part 2", IEEE Multimedia, Volume 7 (1), pages 76-84, 2000.

[2] Sikora T., "The MPEG-7 Visual standard for content description - an overview", IEEE Trans. on Circuits and Systems for Video Technology, special issue on MPEG-7, 11(6):696-702, June 2001.

[3] Coding of Audio-Visual Objects (MPEG-4) – Part 1: Systems, ISO/IEC 14496-1:2001.

[4] ISO/IEC JTC1/SC29/WG11/N3382 14496-1:2001 PDAM2 (MPEG-4 Systems), Singapore, March 2001.

[5] MPEG-21 Overview v.5, ISO/IEC JTC1/SC29/WG11/N5231, Shanghai, October 2002, http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm.

[6] MPEG-4 Overview v.21, ISO/IEC JTC1/SC29/WG11 N4668, March 2002, http://www.chiariglione.org/mpeg/standards/MPEG-4/MPEG-4.htm.

[7] 13. Zhao R. and W.I. Grosky, Narrowing the Semantic Gap-Improved Text-Based Web Document Retrieval Using Visual Features, IEEE Trans. on Multimedia, Special Issue on Multimedia Database, Vol. 4, No 2, June 2002.

[8] E. Papaioannou, K. Karpouzis, P. de Cuetos, V. Karagianis, H. Guillemot, A. Demiris, N. Ioannidis, "MELISA – A Distributed Multimedia System for Multi-Platform Interactive Sports Content Broadcasting", Proceedings of EUROMICRO Conference, pp. 222-229, 2004.

[9] Paliouras G., Papatheodorou C., Karkaletsis V. and Spyropoulos C.D., "Discovering User Communities on the Internet Using Unsupervised Machine Learning Techniques. Interacting with Computers', v. 14, n. 6, pp. 761-791, 2002.

[10] Smith A. S. “Application of Machine Learning Algorithms in Adaptive Web-based Information Systems”,CS-00-01, School of Computing Science Technical Report Series. ISSN 1462-0871, 2001

[11] Miyamoto, S. “Fuzzy Sets in Information Retrieval and Cluster Analysis”. Kluwer Academic Publishers,1990

[12] Duda, R. O., Hart, P. E., & Stork, D. G., “Pattern Classification”, Second Edition, Wiley,2001

[13] Yager R.R., “Intelligent control of the hierarchical agglomerative clustering process”, IEEE Transactions on Systems, Man and Cybernetics, Part B 30(6): 835–845,2000

[14] Wallace, M., Stamou, G. “Towards a Context Aware Mining of Learner Interests for Consumption of Multimedia Documents”, Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), Lausanne, Switzerland,2002

[15] Wallace, M. and Kollias, S., "Soft Attribute Selection for Hierarchical Clustering in High Dimensions", Proceedings of the International Fuzzy Systems Association World Congress (IFSA), Istanbul, Turkey, June-July 2003

[16] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In Proc. of the 10th International World Wide Web Conference (WWW10), Hong Kong, May 2001

[17] Vucetic S. and Obradovic Z., “A Regression-Based Approach for Scaling-Up Personalized Recommender Systems in E-Commerce," Web Mining for E-Commerce Workshop at the Sixth ACM SIGKDD Inl'l Conf. on Knowledge Discovery and Data Mining, Boston, MA., 2000

[18] Hirota, K., Pedrycz, W. (1999) Fuzzy computing for data mining. Proceedings of the IEEE 87:1575–1600

[19] MovieLens, Movie recommendation website, http://movielens.umn.edu

[20] The Internet Movie Database, http://www.imdb.com/

[21] D.M. Pennock and E. Horvitz. Analysis of the axiomatic foundations of collaborative filtering. In AAAI Workshop on Artificial Intelligence for Electronic Commerce, Orlando, Florida, July 1999.

[22] MPEG-21 Overview v.5, ISO/IEC JTC1/SC29/WG11/N5231, Shanghai, October 2002, http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm.

[23] http://www.w3.org/XML/

[24] K. Verbert and E. Duval, “Toward a global architecture for learning objects: A comparative analysis of learning object content models,” in Proc. ED-MEDIA 2004 World Conf. Educational Multimedia, Hypermedia and Telecommunications, 2004, pp. 202–209.

[25] Information Technology-Multimedia Framework (MPEG-21)-Part 2: Digital Item Declaration, ISO/IEC 21000-2:2003, Mar. 2003.

[26] Apache Group, “Apache HTTP Server: Content Negotiation”, http://httpd.apache.org/docs/content-negotiation.html.

[27] K. Holtman and A. Mutz, “Transparent Content Negotiation in HTTP”, IETF RFC 2295, 1998.

[28] K. Munchurl, L. Jeongyeon, K. Kyeongok, and K. Jinwoong, “Agent-Based Intelligent Multimedia Broadcasting within MPEG-21 Multimedia Framework”, ETRI Journal, Volume 26, Number 2, April 2004, pp 136-148.

[29] S. Seshan, M. Stemm, R. H. Katz, “Benefits of transparent content negotiation in HTTP”, Global Internet Mini Conference, Globecom 98, November 1998

[30] O. Steiger, M. Schneider Fontan, D. Marimon Sanjuan, Y. Abdeljaoued, T. Ebrahimi, S. Dominguez, J. San Pedro Wandelmer, N. Denis, F. Granelli and F. De Natale, "Personalized Content Preparation and Delivery for Universal Multimedia Access", No 2005.010, April 2005

[31] S. Kalasapur, M. Kumar, B. Shirazi, "Personalized Service Composition for Ubiquitous Multimedia Delivery", In Proceedings of the Sixth IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks (WoWMoM'05), 2005

[32] J. Masthoff and R. Luckin, "Preface: Workshop Future TV: Adaptive instruction in your living room", In Proceedings of the Future TV: Adaptive Instruction In Your Living Room (A workshop for ITS), 2002

[33] G. Lekakos, K. Chorianopoulos, and D. Spinellis, "Information systems in the living room: A case study of personalized interactive TV design", In Proceedings of the 9th European Conference on Information Systems, June 2001.

[34] Ardissono, L., Kobsa, A., Maybury, M. T. (Eds.), 2004, Personalized Digital Television, Targeting Programs to Individual Viewers, Vol. 6, p. 331

[35] http://www.imdb.com

[36] L. M.J. Hogg, N. R. Jennings, "Socially Intelligent Reasoning for Autonomous Agents", IEEE Transactions on Systems, Man and Cybernetics – Part A , 31(5):381--399, 2001.

[37] Kenneth J. Arrow, "Social Choice and Individual Values", 1951, 2nd ed., 1963, ISBN 0-300-01364-7

[38] M. Papaioannou, K. Pardalis, K. Karpouzis, G. Andreou, N. Ioannidis, "Possible usages of MPEG-21 to improve the MELISA sports broadcasting system", ISO/IEC JTC1/SC29/WG11/M11675, 71st MPEG meeting, Hong Kong, China, January 2005

[39] K. Karpouzis, I. Maglogiannis, E. Papaioannou, D. Vergados, A. Rouskas, "MPEG-21 digital items to support integration of heterogeneous multimedia content", Special issue on Emerging Middleware for Next Generation Networks, Computer Communications, Elsevier, 2006.