The Electronic Road: Personalized Content Browsing

Manolis Wallace, Kostas Karpouzis, George Stamou, George Moschovitis, and Stefanos Kollias
National Technical University of Athens

Christos Schizas
University of Cyprus

Conventional search engines locate terms in queried documents without taking into account their actual semantic content. Search engines also retrieve multimedia documents that aren't necessarily related to each other by semantic continuity. An efficient content-browsing system based on semantic content analysis, user profiling, and adaptation could provide users with more effective content-browsing capabilities.

The interactive infrastructure of the Internet has resulted in a wealth of available multimedia information. In effect, the very nature of the content available on the Internet makes it appealing to different users, depending on their level of expertise, interest, and technological savvy. To locate information on the Internet, most users rely on search engines using text–based searches. Typically, the engine retrieves documents from its index and determines a ranked list of sources to present to the user.[1] However, an obvious disadvantage of this approach is that the search engine must locate the search terms in the queried documents without taking into account their actual semantic content. Another shortcoming is that the retrieved documents aren't necessarily related to each other and therefore represent solitary pieces of information that don't have semantic continuity.

Because users are already familiar with the semantic meaning of the available information, whenever users retrieve a specific information unit, the browsing system could suggest that they continue the retrieval process by viewing semantically related material. As browsing goes on, this process could form an abstract chain of documents that share a common semantic substance—essentially a subset of the available information chosen by the user. You can think of this personal chain of documents as the users' own electronic road through a plethora of semantically related multimedia information.[2]

For example, one electronic road might commence with a document that describes temples in ancient Athens. At this point, the system might suggest different sets of multimedia documents: one might contain general information about ancient Athens while another might consist of information about temples in ancient Greece. When users choose a new document from one of these sets, the system enriches the particular electronic road with a new node—a new stop on the highway of multimedia documents that share common subjects with the user's previous choices. That electronic road might continue with nodes describing temple structures in the greater Mediterranean area or even temples of the area spanning different chronological epochs.[3]

From the system's point of view, an electronic road represents the user's navigation path through the multimedia content. Specifically, it represents the series of links to the system's Information Units (IUs) that the user follows. IUs are the building blocks of the available multimedia content and consist of the actual data—such as video segments, images, sound clips, or text documents—along with the attached metadata. While displaying the requested IU, the system also produces several dynamic links that point the user to related IUs. These links extend the current electronic road according to the IU semantic nature and the user's profile.

In addition to storing the particular user's personal experience, it's possible to use this sequential chain of documents in different ways. For example, the common semantic context of this electronic road could indicate these stops as typical of a specific subject, such as the ancient temples from the previously noted example. An integrated system could suggest this specific road to another user interested in the same subject. In most cases, content-browsing systems suggest individual multimedia documents related to the one that the user is currently viewing. Extending the suggestions to a complete electronic road could enhance the notion of the semantic context being shared.

1070-986X/03/$17.00 Published by the IEEE Computer Society

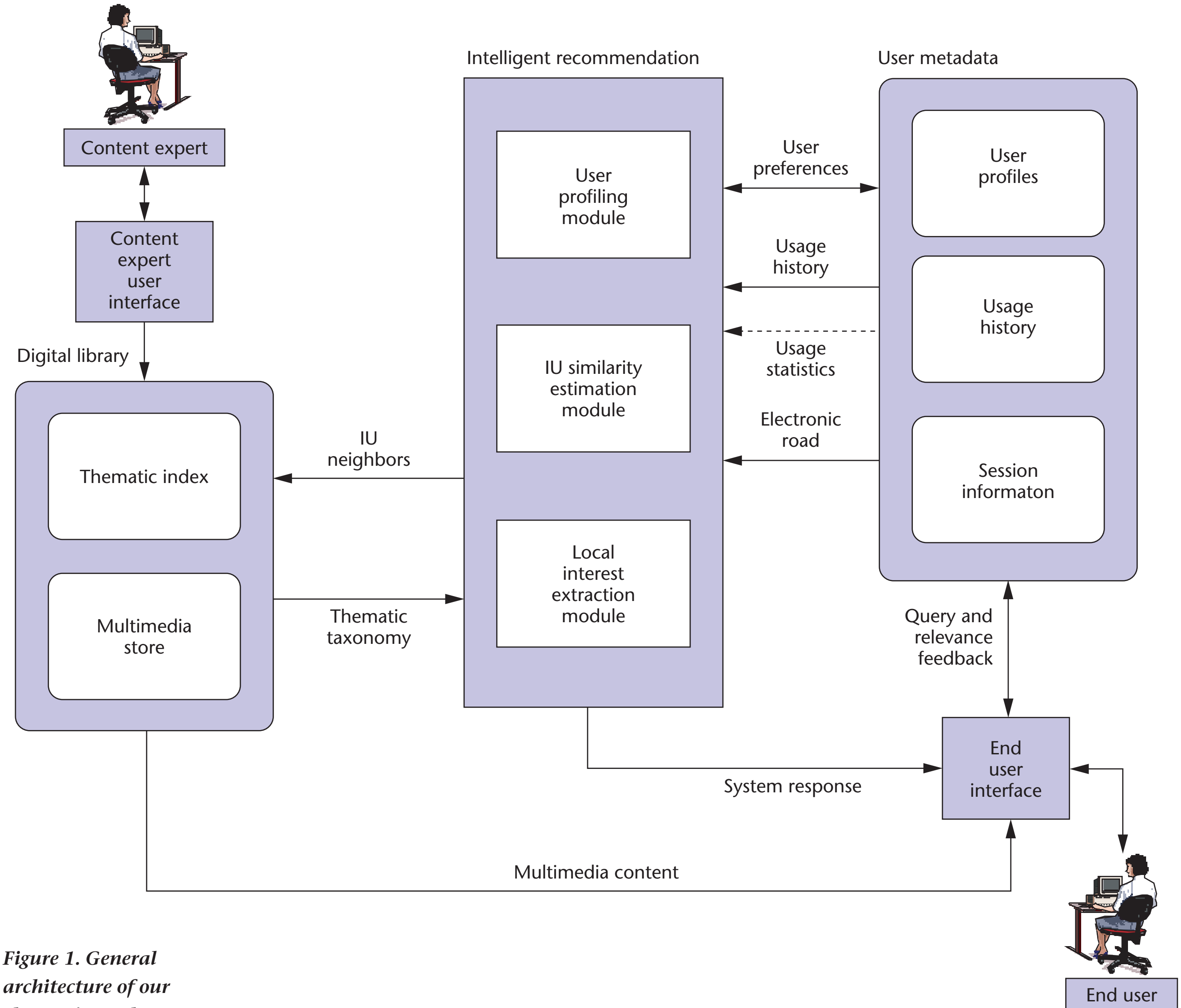

Figure 1. General architecture of our electronic road system.

System architecture

Figure 1 defines the modules and the flow of information in our proposed system. The information stored in the system includes both the multimedia material and the metadata used by the recommendation modules to provide the user with context-based browsing.

The multimedia information is stored in the digital library, which includes two components: the multimedia store and the thematic index. The multimedia store contains multimedia material in several distinct types, including images, video, audio, and text in the form of multimedia documents. The thematic index contains the metadata information provided by the content expert user for each document. This metadata information, which relates documents to the predefined thematic categories, comprises the distinct IUs. The thematic index also contains relations associating the IUs to each other.

Traditionally, information-retrieval systems model the user as an entity that has a set of interests; a portion of the system's IUs correspond to these interests. Through various methods, the user attempts to interact with the system to retrieve the documents that correspond these interests. In our system, the key point is determining user interest and filtering the multimedia information according to those interests. To determine the user interest at any time for any specific user, we must address three kinds of user information to generate the user metadata: the usage history, the session information, and the user profile.

The usage history tracks all past user transactions, while the session information describes all the transactions the user made in the current session. We determine the user profile, which contains the user preferences, from the history of transactions as well as from the user's static profile. We represent the user profile through a set of vectors that describe the user's preferences with respect to the information retrieval attributes. We can map this information onto a neural network to adopt the system's behavior to the user's profile and the current user interest.

The system uses all of this information for the purpose of making intelligent recommendations. Three separate modules work to make these intelligent recommendations. The *IU similarity estimation module* operates on the thematic index and enriches it by including estimations of similarities among IUs. The *user profiling module* extracts the user preferences out of the usage history and stores them in the user profiles. The system uses this information to personalize the offered information services. Finally, the *local interest extraction module* analyzes the session information to determine the context provided by both the user profile and the electronic road.

IU similarity estimation module

The system's goal is to help users browse multimedia IUs. To do so, the system must be able to propose to the user IUs that are related to the user's current preferences, which are defined by the user's profile and by the IUs the user has recently shown interest in. In a full-scale retrieval system, where the total number of IUs is huge, it's not efficient to apply algorithms that rely on comparisons between some IUs of interest and all remaining IUs in the system. The document space of interest should be limited to make real-time operation feasible. In our system, we loosely precalculate IU-to-IU similarities and store them in the index. We can use the index to estimate all related IUs quickly.

Our system describes each IU using a predefined set of features, including those related to theme (such as degree of relevance to politics), location (such as degree of relevance to Greece), and so forth. The system could automatically characterize the IUs using the techniques described elsewhere.[4-7] On the basis of each one of these features, we define a distance measure between IUs as $Dist_i(IU_1, IU_2) = Dist_i(\mathbf{x}_1, \mathbf{x}_2) = 1 - \min(\mathbf{x}_1^i, \mathbf{x}_2^i)$ where $Dist_i(IU_1, IU_2)$ is the distance between IU_1 and IU_2 when considering feature i, and $\mathbf{x}_1^i$, $\mathbf{x}_2^i$ are features i of the IUs—$\mathbf{x}_a$ is the vector associated with information unit IU_a.

In offline mode, before the user inserts a query, the system can't foresee the topics of subsequent user interaction. Therefore, at this stage it's only possible to estimate IU distances without any consideration of the context. Thus, we can consider two IUs too similar if any context exists through which their distance is small. Formally, we consider two IUs to be generally similar to the degree that

$$Sim(\mathbf{x}_1, \mathbf{x}_2) = 1 - \frac{\sum_{i=1}^{F} [t_i]^{\lambda} Dist_i(\mathbf{x}_1,}{\sum_{i=1}^{F} [t_i]^{\lambda}}$$

is small, for some weighting t of features, where λ is a predefined constant and F is the count of features. In the most typical case, when $\lambda \neq 1$ and $Dist_i(\mathbf{x}_1, \mathbf{x}_2) \neq 0$ for every $i \in 1 \ldots F$, then we can prove, by demanding that the derivative of $Sim(\mathbf{x}_1, \mathbf{x}_2)$ with respect to each feature $i \in 1 \ldots F$ is zero, that the optimal weighting t is given as

$$t_i = t_1 \left[\frac{Dist(\mathbf{x}_1^1, \mathbf{x}_2^1)}{Dist(\mathbf{x}_1^i, \mathbf{x}_2^i)} \right]^{\frac{1}{\lambda - 1}}, i \in 2 \ldots F$$

and

$$t_1 = \frac{1}{\sum_{i=1}^{F} \left[\frac{Dist(IU_1^1, IU_2^1)}{Dist(IU_1^i, IU_2^i)} \right]^{\frac{1}{\lambda - 1}}}$$

In the first of the two remaining special cases, when $\lambda = 1$, it's easy to see that the weighting that produces the best similarity is the one that promotes the features for which the distance among the two IUs is smallest. In the second case, when $\lambda \neq 1$, if features exist for which the distance among the two IUs is zero, then the optimal weighting is the one that promotes exactly those features.

Because the system calculates the optimal weighting of the features analytically, estimating the similarity between two IUs has the ideal complexity of $O(1)$. Thus, the process of comparing all IUs to each other has a polynomial complexity of $O(n^2)$, as exactly $n(n-1)/2$ comparisons are needed, where n is the count of IUs in the system and we can easily apply it in offline mode.

User profiling module

The most important information the system uses is the electronic road information. However, when the user initiates a new session, an electronic road has not yet been formed, or the road is too short to provide sufficient information concerning the user's quest. The system must have access to an alternative form of information concerning the user's possible quests, at least until the electronic road has been sufficiently initialized. We call this information the *user profile module*.

Usage history analysis

For each user, all IUs are initially noted as indifferent and stored in the usage history. The system keeps only a small subset of them for the sake of both space and processing time in subsequent algorithms. When the user interacts with the system, all information is stored as usage history, which consists of the IUs for which the user has indicated interest and the IUs for which the user has indicated dislike. The user indicates interest when first previewing an IU and then deciding to use it to continue the journey down the electronic road using similar IUs.

Users state dislike, on the other hand, only explicitly by marking the IU as not interesting. The user might use a dislike selection to refine the context of the browsing session, but this selection might not necessarily correspond to a general dislike. Therefore, if both an interest and a dislike exist in the usage history for a specific IU, the system removes the latter. Likewise, if positive or negative interests appear for any IUs that also exist in the set of indifferent IUs, or the index shows that they are similar to IUs in the set of indifferent ones, then the system removes the latter. In short, interests are most reliable, followed by dislikes and indifferent IUs.

The system analyzes usage history with the aid of a context-aware hierarchical clustering algorithm. This algorithm can identify the count and characteristics of distinct patterns in the set of interesting, disliked, or indifferent IUs in a user's usage history. For each one of these sets, every IU is initially turned into a singleton—a cluster of one element. Then the system merges the two most similar clusters recursively until no more similar clusters remain. We can describe each one of the resulting clusters, which corresponds to a distinct pattern in the usage history, by its center and standard deviations in the F-dimensional space.

The key to this algorithm is to meaningfully estimate distances among any two clusters. We base the intercluster distance on the notion of context developed in the previous section. Thus we define each feature i, a distinct distance among clusters, as

$$Dist_i(c_1, c_2) = \sqrt[\kappa]{\frac{\sum\limits_{\mathbf{x}_a \in c_1, \mathbf{x}_b \in c_2} \left[Dist_i(\mathbf{x}_a, \mathbf{x}_b\right.}{|c_1||c_2|}}$$

In this equation, κ is a constant. We use $\kappa = 2$, because the system can compute powers of 2 and 0.5 much more efficiently than others in existing computer systems. Then we can easily estimate the overall distance among the two clusters as

$$Dist(c_1, c_2) = \frac{\sum\limits_{i=1}^{F} [t_i]^{\lambda} Dist_i(c_1, c_2)}{\sum\limits_{i=1}^{F} [t_i]^{\lambda}}$$

We can easily calculate this equation once the system finds the optimal context t using the methodology developed in the previous section.

We base the termination of the clustering on a predefined threshold on the value of intercluster distances. Because we estimate the context for the comparison of two clusters with an $O(1)$ algorithm, we can keep the complexity of the proposed context-aware approach to the same small polynomial levels as traditional Euclidian-based hierarchical clustering. Together with the fact that the extremely large set of indifferent IUs is limited with the use of principal components analysis, this technique makes our approach applicable in offline mode in full-scale applications.

Mapping the user profile

We use the resulting clusters to initialize a three-layer neural network classifier that contains all the personalization information available in the usage history. The system uses the input layer, which has F input nodes, to collect features for any given document. As Figure 2 shows, the hidden layer has one radial basis function (RBF) node for each cluster c that the analysis of the usage history generated. These RBF nodes are specified by their centers and spreads in the F-dimensional space, initialized as centers $\mathbf{c}_{\mu}$ and standard deviations $\mathbf{c}_\sigma$ of the corresponding clusters.[8] We can use different types of output node functions.

We select the sigmoid as a nonlinear discriminant function, linking each one of the hidden nodes to one of the output nodes, depending on the type of IUs the corresponding cluster contains. Adapting the neural network to the context of a formed electronic road can alter all of its parameters $\mathbf{w}$—that is, the values of $\mathbf{c}_\mu$, $\mathbf{c}_\sigma$, and $\mathbf{t}$.

Given the way the system creates the neural classifier, it's easy to see that when it's fed with a document, the activation of its output layer will show whether usage history indicates that the user has an interest in the document, dislikes it, or is indifferent to it. Additionally, the system uses the neural network to determine when to analyze usage history for the extraction of user profiles.

When a user completes a session, the system adds the electronic road information to the usage history, which consists of a set S_c of IUs and corresponding classifications that consist of m_c pairs $S_c = \{(\mathbf{x}_i, \mathbf{d}_i), \ldots, (\mathbf{x}_{mc}, \mathbf{d}_{mc})\}$, where $\mathbf{x}_i$ and $\mathbf{d}_i$, with $i = 1, 2, \ldots, m_c$, similarly correspond to input i and classification data. Each $\mathbf{x}_1$ of these IUs is fed to the user's neural network. If IUs exist for which the output $\mathbf{y}(\mathbf{x}_i)$ of the network is different from the desired classification $\mathbf{d}_i$, then the system rebuilds the user profile, which implies that the neural network parameters aren't compatible with the newly acquired knowledge.

Extraction of the local interest

Whenever a user views an IU, the system considers the user to be on an electronic road. While the user interacts with the system and forms an electronic road, the proposed system analyzes the IUs for which he or she has provided positive relevance feedback; estimates the context; and proposes to him or her the IUs that are most suitable for the continuation of this electronic road. The system accesses the user profile to initiate the electronic road context, but only considers IUs in the current IU neighborhood, as specified in the index, relating each IU to all other IUs.

Local interest detection

The system expresses each IU, as well as each user profile, as an attribute vector, relating the IU or user with a predefined number of attributes, including thematic information and other characteristics. The system forms user profiles on the basis of appropriate usage-history clustering and maps them onto radial-basis-function neural networks. As a result, the system assumes that a network is available for each user. The system detects a user's local interests when forming an electronic road by adapting the neural network parameters that correspond to the general user's profile so that the network can match the specific user's selections and rejections with the presented IUs.

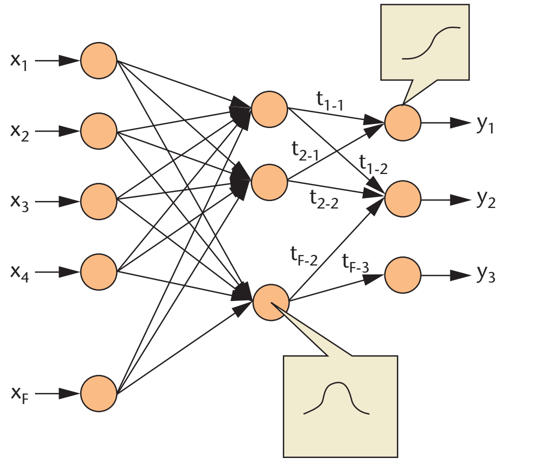

Figure 2. The radial basis function (RBF) neural network structure.

In particular, the network classifies each IU attribute vector, $\mathbf{x}_i$, to one of the three classes ω_1, ω_2, or ω_3 of the three output states (interesting, indifferent, or disliked). The system defines the output vector $\mathbf{y}(\mathbf{x}_i) = (p^i_{\omega_1}\ p^i_{\omega_2} \ldots p^i_{\omega_p})^T$, where $p^i_{\omega_j}$ denotes the probability that the network input i belongs to class j.

For each user, the system keeps in the usage history a set of IUs that the user has already selected as $S_b = \{(\mathbf{x}'_i, \mathbf{d}'_i), \ldots, (\mathbf{x}'_{mb}, \mathbf{d}'_{mb})\}$, where vectors $\mathbf{x}'_i$ and $\mathbf{d}'_i$ with $i = 1, 2, \ldots, m_b$ denote the input i and corresponding desired output vectors. The network should have the ability to adapt its performance according to this information and the formation of the user's electronic road.

To consider network adaptation in more detail, vector $\mathbf{w}_b$ can include the network parameters before adaptation, and $\mathbf{w}_a$ is the new parameter vector obtained through adaptation. We assume a retraining set, S_c, to be extracted from the electronic road that the user forms. This set consists of, for example, m_c feature vectors $S_c = \{(\mathbf{x}'_1, \mathbf{d}'_1), \ldots, (\mathbf{x}'_{mc}, \mathbf{d}'_{mc})\}$. The system performs the adaptation using efficient network training and computes new network parameters $\mathbf{w}_a$ by minimizing the following error criterion with respect to the parameters

$$E_a = E_{c,a} + \eta E_{f,a}$$

$$E_{c,a} = \frac{1}{2}\sum_{i=1}^{m_c} \left\| \mathbf{y}_a(\mathbf{x}_i) - \mathbf{d}_i \right\|_2$$

$$E_{f,a} = \frac{1}{2}\sum_{i=1}^{m_b} \left\| \mathbf{y}_a(\mathbf{x}'_i) - \mathbf{d}'_i \right\|_2$$

In this equation, $E_{c,a}$ is the error performed over the retraining set S_c (the current electronic road), and $E_{f,a}$ is the corresponding error over the original training set S_b (the former knowledge). Also, $\mathbf{y}_a(\mathbf{x}_i)$ and $\mathbf{y}_a(\mathbf{x}'_i)$, which correspond to input vectors $\mathbf{x}_i$ and $\mathbf{x}'_i$, are the outputs $\mathbf{y}(\mathbf{x}_i)$ and $\mathbf{y}(\mathbf{x}'_i)$ of the adapted network that consists of parameters $\mathbf{w}_a$. Similarly, $\mathbf{y}_b(\mathbf{x}_i)$ represents the output of the network that consists of parameters $\mathbf{w}_b$ when accepting vector $\mathbf{x}_i$ at the input. Parameter η is a weighting factor accounting for the significance of the current retraining set compared to the former one and $\|\cdot\|_2$ denotes the L_2-norm.

Each time the system requires adaptation, it creates a new set S_c that represents the current condition. Then the system estimates new network parameters, taking into account both the current information (the data in S_c) and the former knowledge (the data in S_b). Because the set S_c is collected within the current electronic road, the system sequentially augments it to include all user selections in the same electronic road and transfers it to usage history at the end of the session.

The adaptation procedure's goal is to minimize E_a and estimate the new network parameters $\mathbf{w}_a$. We first assume that a small perturbation of the network parameters $\mathbf{w}_b$ is enough to achieve good classification performance. Then $\mathbf{w}_a = \mathbf{w}_b + \Delta\mathbf{w}$, where $\Delta\mathbf{w}$ is a small incremental vector. This assumption leads to a tractable solution for estimating $\mathbf{w}_a$, because it permits linearization of the nonlinear activation function of the network output neurons using a first-order Taylor series expansion.

Defining E_a indicates that the new network parameters take into account both the current and the previous network information. However, you can replace the definition of $E_{c,a}$ with the constraint that the actual network outputs are equal to the desired ones. That is, $\mathbf{y}_b(\mathbf{x}_i) = d_i$, i = 1 … , m_c for all data in S_c, which indicates that the first component of E_a, corresponding to error $E_{c,a}$, takes values close to zero after estimating the new network parameters.

It's possible to show[9] that, through linearization, satisfying this constraint with respect to the weight increments is equivalent to a set of linear equations $\Lambda = \boldsymbol{A} \cdot \Delta\mathbf{w}$, where vector Λ and matrix $\boldsymbol{A}$ are appropriately expressed in terms of the previous network parameters. In particular, $\Lambda=[\mathbf{y}_a(\mathbf{x}_1)\ldots\mathbf{y}_a(\mathbf{x}_{m_c})]^T - [\mathbf{y}_b(\mathbf{x}_1)\ldots\mathbf{y}_b(\mathbf{x}_{m_c})]^T$, expressing the difference between network outputs after and before adaptation for all input vectors in S_c. Applying the constraint, we can write vector Λ as $\Lambda=[d_1\ldots d_{m_c}]^T - [\mathbf{y}_b(\mathbf{x}_1)\ldots\mathbf{y}_b(\mathbf{x}_{m_c})]^T$.

The size of vector Λ is smaller than the number of unknown parameters $\Delta\mathbf{w}$, because, in general, a small number, m_c, of IUs forms the electronic road. Thus many solutions exist for the set of linear equations because the number of unknowns is much greater than the respective number of equations. However, an additional requirement imposes uniqueness, which takes into account the previous network information. Among all possible solutions, the system will select one that causes a minimal degradation of the previous network information. However, the system estimates the network parameters before adaptation, $\mathbf{w}_b$, as an optimal solution over data of set S_b.

Furthermore, the parameters after adaptation provide a minimal error over all data of the current set S_c. Thus, minimizing $E_{f,a}$, which expresses the effect of the new network parameters over data set S_b, is equivalent to minimizing the absolute difference of the error over data in S_b with respect to the previous and the current network parameters. This means that the parameter increments are minimally modified, resulting in the following error criterion $E_s = \|E_{f,a} - E_{f,b}\|_2$, with $E_{f,b}$ defined similarly to $E_{f,a}$, and $\mathbf{y}_a$ replaced by $\mathbf{y}_b$.

We can show that this process will take the form of $E_S = 1/2(\Delta\mathbf{w})^T \cdot \boldsymbol{K}^T \cdot \boldsymbol{K} \cdot \Delta\mathbf{w}$, where the elements of matrix $\boldsymbol{K}$ are expressed in terms of the previous network parameters $\mathbf{w}_b$ and the training data in S_b. The problem then results in minimizing E_S. We don't encounter overfitting problems in adaptation because the training data—that is, the number of data in S_c and S_b—is greater than the respective network parameters.

The error function E_S is convex because it's squared, while the constraints are linear equalities. Thus the solution should lie on the hypersurface defined by $\Lambda = \boldsymbol{A} \cdot \Delta\mathbf{w}$ and simultaneously minimize E_S. We use the gradient-projection method to solve this problem. The philosophy of the gradient-projection method is to move in a direction that decreases E_S and simultaneously satisfies the constraints. A point is feasible if it satisfies all constraints. We adapt the parameter increments as $\Delta\mathbf{w}(n + 1) = \Delta\mathbf{w}(n) + \mu(n)\mathbf{h}(n)$, where $\mu(n)$ is a scalar that determines the rate of

Table 1. Features for document descriptions and profile representation. The first three lines list thematic categories, the fourth line years, and the fifth line location.

Sports	Arrivals–Departures	Industry–Commerce	Transportation	Celebrations	Ecclesiastical Themes
Government	Public Services	Artistic	Politics	Naval	Tourism
Historical Events	Head of State	Military	Celebrities	—	—
1910s	1920s	1930s	1940s	1950s	1960s
Greece	Turkey	Cyprus	Balkans	—	—

convergence. We can estimate vector $\mathbf{h}(n)$ as

$$\mathbf{h}(n) = -\mathbf{P}\nabla E_s = -\mathbf{Q}\Delta\mathbf{w}$$
$$\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{A}^T(\boldsymbol{A}\,\boldsymbol{A}^T)^{-1}\boldsymbol{A}$$
$$\boldsymbol{Q} = \boldsymbol{P}\boldsymbol{K}^T\mathbf{K}\Delta\mathbf{w}$$

In this equation, we use ∇E_S computed from E_S=1/2$(\Delta\mathbf{w})^T \cdot \boldsymbol{K}^T \cdot \boldsymbol{K} \cdot \Delta\mathbf{w}$ at iteration n. The computational complexity required to update each network parameter independently is low, as it's linear with respect to the number of network parameters.

IU recommendation

More traditional content-browsing systems don't rely on user selection to determine the context of user interaction. Thus, supposing that the user profile isn't altered, the system always proposes the same documents to the user when the user selects a specific category. In some systems, because numerous documents are related to each thematic category, certain documents will never be displayed to a user through content browsing. Our approach overcomes this problem as it moves through documents in a browsing session using the estimated context of the session together with the user profile to filter available documents.

Our approach is related to relevance feedback in information-retrieval systems. The fundamental difference between these systems and the one presented here is that they're designed for operation in a search mode; they don't incorporate relevance feedback in a browsing mode. Thus, in these systems, a user might need to know of the existence of a document or subject to locate it through a combination of initial query and feedback interactions.

We based the operation of our proposed system on neural network structures to incorporate the context of the electronic road into the recommendation process. It's not possible to display all of the documents that are related to a thematic category, which means that the system must perform some sort of document selection. Because no more feedback is yet available from the user, the only available information is in the user profile, so the system must first consult the thematic index to acquire the set of documents related to the category chosen by the user. Out of those documents, the system recommends the ones most related to the user profile.

When the set of recommended documents is available, the user has several options. First, the user can view one of the IUs, which the system will consider to be an indication of interest. The user could also mark one of the IUs as not interesting, which the system will consider to be an indication of dislike. The user can also decide to continue his or her path on the electronic road. In this case, the system must produce a new set of recommended documents. Prior to doing so, however, the new session information must be incorporated in the neural network. The system will add the last user choices in set S_c to retrain the neural network. Following this, the system generates a set of documents to display to the user.

Experimental results

We implemented the system in an experimental framework as part of the Cultural Journeys in the Information Society (CJIS) project. For the purposes of the experiment, we populated a database with 10,000 videos from the archive of the Greek Ministry of Press and Mass Media. These videos contained diverse footage, dating from the early 1900s, annotated by an expert historian. We used these annotations to decide on the relevance of each video with respect to each of the 16 predefined thematic categories shown in Table 1. From these scores, the system calculated IU-to-IU distances offline. We used these distances, stored in the index, to make document neighborhoods readily available.

In the following example, a user browses documents in the thematic category Military. Initially, the neural network is based on the user profile, the specific weights of which we show in

Table 2. The preference weights from a user profile mapped to the themes shown in Table 1.

0.4, 0.3	0.8, 0.15	0.3, 0.4	0.75, 0.25	0.8, 0.15	0.3, 0.4
0.7, 0.3	0.4, 0.4	0.4, 0.4	0.7, 0.3	0.4, 0.3	0.3, 0.4
0.95, 0.05	0.9, 0.1	0.65, 0.3	0.4, 0.4	—	—
0.9, 0.05	0.8, 0.1	0.75, 0.1	0.65, 0.3	0.65, 0.3	0.65, 0.3
0.9, 0.05	0.8, 0.35	0.8, 0.35	0.7, 0.4	—	—

(a)

(b)

(c)

(d)

(e)

Figure 3. Initial suggestion of documents related to the Military thematic category: (a) A selection of video shots showing soldiers marching in the countryside; (b) video shots from the port of Piraeus; (c) two shots from an army gathering; (d) video shots from an army and civilian gathering near the city entrance, next to the lake; and (e) sailors and senior officers parading in front of officials not shown in the footage.

Table 3. The adjusted preference from the user profile mapped to the themes shown in Table 1.

0.4, 0.3	0.7, 0.25	0.3, 0.4	0.75, 0.35	0.4, 0.35	0.2, 0.4
0.6, 0.3	0.3, 0.4	0.3, 0.4	0.7, 0.25	0.95, 0.2	0.1, 0.4
0.90, 0.15	0.7, 0.2	0.7, 0.25	0.35, 0.4	—	—
0.6, 0.25	0.95, 0.05	0.55, 0.3	0.4, 0.4	0.45, 0.35	0.45, 0.4
0.9, 0.05	0.6, 0.4	0.55, 0.45	0.7, 0.3	—	—

Table 2. The retrieval system, using the profile-generated neural network to filter documents in the index, suggests the IUs shown in Figure 3 along with short summaries of their description, sorted by relevance.

The user selects the document shown in Figure 3b, effectively narrowing the session context to documents related to the Greek navy. The system performs the neural network adaptation procedure, taking into account this selection. Following this, the system presents the user with a new set of documents, shown in Figure 4. On the basis of the new selections shown in Figure 4, the neural network provides adjusted parameters for the hidden layer node whose initial parameter values are presented in Table 2. Table 3 shows the updated weights.

The network structure now describes a session context that's no longer about the military in general, but about the navy in particular. As a result, the next suggestion by the system is the document shown in Figure 5. The context identified by the system includes IUs closely related to the thematic categories Military and Navy. With respect to their actual content, the system ranks IUs in Figure 6 higher than the rest because their date is closer to that of the session context.

Conclusions

The system we presented—a content-browsing system based on the electronic road metaphor—could help users browse multimedia documents by suggesting material related to what users are viewing. Our proposed system could achieve this goal by combining robust information clustering with adaptive neural-network-based interest updates. In particular, the system could cluster each user's usage history, generating a user profile to initialize user browsing through multimedia content. This initialization could include an automatic mapping of the user profile onto an adaptive neural network to detect user interests.

Extending this exploration in the framework of content-based multimedia analysis, as well as MPEG-7 and MPEG-21 standardization activities, is a topic of our current research and development. In the framework of the EU FP5 IST-1999-20502 Faethon project, to be completed in January 2004, the group has developed a software prototype that relies on the principles presented here and will allow for intelligent, personalized retrieval of MPEG-7 content. Heterogeneous, non-MPEG-7 archives will be supported through developing custom archive interfaces to the system. The prototype system will be installed in the

(a) **(b)**

Figure 4. Other documents the user selected: (a) a corps of sailors and senior officers parading in front of officials not shown in the footage and (b) the Turkish Navy, with the flagship Yavouz at head, sails in the Faliro Bay.

Figure 5. The system recommendation after the first formation of the electronic road: a selection of video shots of fleet admiral Kountouriotis, aboard the warship Averof, next to ship commander Admiral Dousmanis.

(a) **(b)** **(c)**

Figure 6. The system recommendation, when the second electronic road has been formed: (a) The funeral of King George I, murdered by Alexandros Schinas in Salonika, and transportation of his corpse from the Athens cathedral to the train station; (b) celebration of the second anniversary of the August coup, held in Athens on 4 August 1938; and (c) video shots from an army and civilian gathering near the city entrance next to the lake.

Greek national TV archive (ERT) and in the Film Archive Austria (FAA), while also being evaluated by Alinari (Italy) and Catalunya TV.

Moreover, within an EU FP6 four-year project called Acemedia, the group will investigate interweaving these notions with semantic-knowledge technologies, developing intelligent adaptive semantic knowledge-based systems for real-life multimedia applications. In addition, the group will collaborate in a similar context with the major European semantic Web groups. **MM**

Acknowledgments

We thank the Ministry of Press and Mass Media for their kind permission to use the images for this article. This work has been partially funded by the EU IST-1999-20502 Faethon project, the INCO Cultural Journeys in the Information Society (CJIS) project, and the Greek National Project "Digitisation & Access to the Film Archive of the Ministry of Press and Mass Media."

References

1. R.A. Baeza-Yates and B.A. Ribeiro-Neto, *Modern Information Retrieval*, Addison-Wesley, 1999.
2. C. Constantinou et al., "Constructing Electronic Roads in the Information Society: A Case Study in Cyprus," *Proc. 4th Int'l Conf. Networking Entities*, 1999.
3. M. Wallace et al., "Towards Supporting the Teaching of History Using an Intelligent Information System That Relies on the Electronic Road Metaphor," *Proc. 3rd IEEE Int'l Conf. on Advanced Learning Technologies*, IEEE CS Press, 2003, pp. 47-51.
4. N. Dimitrova et al., "Applications of Video-Content Analysis and Retrieval," *IEEE MultiMedia*, vol. 9, no. 3, 2002, pp. 42-55.
5. G. Iyengar and A. Lippman, "Models for Automatic Classification of Video Sequences," *Proc. SPIE Conf. Storage Retrieval Still Image Video Databases*, International Society for Optical Engineering, 1998, pp. 216-227.
6. M.R. Naphade and T.S. Huang, "Extracting Semantics from Audio-Visual Content: The Final Frontier in Multimedia Retrieval," *IEEE Trans. Neural Networks*, vol. 13, no. 4, 2002, pp. 793-810.
7. M. Wallace, G. Akrivas, and G. Stamou, "Automatic Thematic Categorization of Documents Using a Fuzzy Taxonomy and Fuzzy Hierarchical Clustering," *Proc. IEEE Int'l Conf. Fuzzy Systems*, IEEE CS Press, 2003, pp. 1446-1451.
8. N. Tsapatsoulis, M. Wallace, and S. Kasderidis, "Improving the Performance of Resource Allocation Networks through Hierarchical Clustering of High-Dimensional Data," *Proc. Int'l Conf. Artificial Neural Networks*, LNCS 2714, Springer, June 2003, pp. 867-874.
9. N. Doulamis, A. Doulamis, and S. Kollias, "Online Retrainable Neural Networks: Improving Performance of Neural Networks in Image Analysis," *IEEE Trans. on Neural Networks*, vol. 11, no. 1, Jan. 2000, pp.1-20.

Manolis Wallace is an assistant professor of computer science at the University of Indianapolis, Athens Campus. His research interests include data mining and intelligent information retrieval. Wallace is pursuing a PhD at the National Technical University of Athens.

Kostas Karpouzis is a senior researcher at the Image, Video and Multimedia Systems Laboratory at the National Technical University of Athens. His research interests include human–computer interaction, image and video processing, human modeling and animation, and virtual reality. Karpouzis received PhD in face animation from the National Technical University of Athens.

George Stamou is a senior researcher at the Institute for Communication and Computer Systems at the National Technical University of Athens. His research interests include fuzzy sets, hybrid intelligent systems, and human–computer interaction. Stamou received a PhD in computational intelligence from the National Technical University of Athens.

George Moschovitis is a researcher at the Image, Video and Multimedia Systems Laboratory at the National Technical University of Athens. His research interests include software development and information retrieval. Moschovitis is pursuing a PhD at the National Technical University of Athens.

Stefanos Kollias is a professor at the National Technical University of Athens. His research interests include image and video processing, analysis, coding, storage, retrieval, multimedia systems, computer graphics, virtual reality, artificial intelligence, neural networks, human–computer interaction, and medical imaging. Kollias received a PhD in signal processing at the National Technical University of Athens.

Christos Schizas is a professor at the University of Cyprus. His research interests include computational intelligence, artificial neural networks, systems theory, applications in medicine, and diagnostic systems. Schizas received a PhD in systems theory at the Queen Mary College, University of London.

Contact Stefanos Kollias at the Image, Video, and Multimedia Systems Lab, Electrical and Computer Engineering Dept., Nat'l Technical Univ. of Athens, Politechnioupoli, 157 73, Zografou, Greece; stefanos@cs.ntua.gr.

Correction

In the July–September 2003 issue, the article "Internet Delivery of MPEG-4 Object-Based Multimedia" by Yasser Pourmohammadi-Fallah, Kambiz Asrar-Haghighi, and Hussein Alnuweiri showed the three boxes in Figure 2, p. 70, labeled as "Compression layer." The correct version is reprinted here. *IEEE MultiMedia* regrets this error.

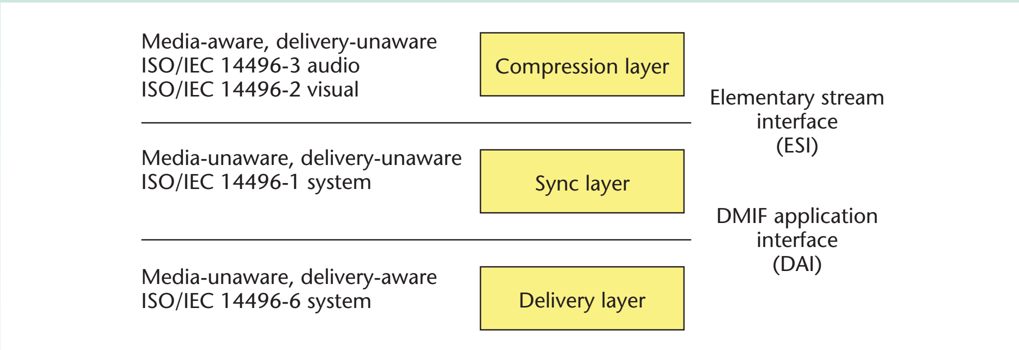

Figure 2. MPEG-4 terminal architecture.

Correction

In the October–December 2003 issue—in the article "The Electronic Road: Personalized Content Browsing" by Manolis Wallace, Kostas Karpouzis, George Stamou, George Moschovitis, Stefanos Kollias, and Christos Schizas—an error occurred with some of the equations not printing fully on the page. We're reprinting these equations here. *IEEE MultiMedia* regrets this error.

$$Sim(\mathbf{x}_1, \mathbf{x}_2) = 1 - \frac{\sum_{i=1}^{F} [t_i]^{\lambda} Dist_i(\mathbf{x}_1, \mathbf{x}_2)}{\sum_{i=1}^{F} [t_i]^{\lambda}}$$

page 51

$$Dist_i(c_1, c_2) = \sqrt[\kappa]{\frac{\sum_{\mathbf{x}_a \in c_1, \mathbf{x}_b \in c_2} [Dist_i(\mathbf{x}_a, \mathbf{x}_b)]^{\kappa}}{|c_1||c_2|}}$$

page 52

$$E_{c,a} = \frac{1}{2} \sum_{i=1}^{m_c} \| \mathbf{y}_a(\mathbf{x}_i) - \mathbf{d}_i \|_2$$

page 54

$$E_{f,a} = \frac{1}{2} \sum_{i=1}^{m_b} \| \mathbf{y}_a(\mathbf{x}_i') - \mathbf{d}_i' \|_2$$

page 54