International Journal on Artificial Intelligence Tools
Vol. XX, No. X (2014) 1–21

TOWARDS IMPLICIT USER MODELING BASED ON ARTIFICIAL INTELLIGENCE, COGNITIVE STYLES AND WEB INTERACTION DATA ‡

EFI PAPATHEOCHAROUS

Department of Computer Science, University of Cyprus
P.O. Box 20537, 1678 Nicosia, Cyprus
efi.papatheocharous@cs.ucy.ac.cy
and
SICS Swedish ICT AB
Isafjordsgatan 22/Kistagången 16, Box 1263
SE-164 29 Kista, Stockholm, Sweden
efi.papatheocharous@sics.se

MARIOS BELK

Department of Computer Science, University of Cyprus
P.O. Box 20537, 1678 Nicosia, Cyprus
belk@cs.ucy.ac.cy

PANAGIOTIS GERMANAKOS

Department of Computer Science, University of Cyprus
P.O. Box 20537, 1678 Nicosia, Cyprus
and
SAP AG, Dietmar-Hopp-Allee 16, 69190 Walldorf, Germany
pgerman@cs.ucy.ac.cy

GEORGE SAMARAS

Department of Computer Science, University of Cyprus
P.O. Box 20537, 1678 Nicosia, Cyprus
cssamara@cs.ucy.ac.cy

A key challenge of adaptive interactive systems is to provide a positive user experience by extracting implicitly the users' unique characteristics through their interactions with the system, and dynamically adapting and personalizing the system's content presentation and functionality. Among the different dimensions of individual differences that could be considered, this work utilizes the cognitive styles of users as determinant factors for personalization. The overarching goal of this paper is to increase our understanding about the effect of cognitive styles of users on their navigation behavior and content representation preference. We propose a Web-based tool, utilizing Artificial Intelligence techniques, to implicitly capture and find any possible relations between the cognitive styles of users and their characteristics in navigation behavior and content representation preference by using their Web interaction data. The proposed tool has been evaluated with a user study revealing that cognitive styles of users have an effect on their navigation behavior and content representation preference. Research works like the reported one are useful for improving implicit and intelligent user modeling in engineering adaptive interactive systems.

‡ A preliminary, condensed version of this work was presented in Mining Humanistic Data Workshop 2012 [35]

Keywords: User Modeling; Artificial Intelligence; Cognitive Styles; Web Engineering; Personalization; Clustering; Sequence Vectors.

1. Introduction

Engineering interactive systems under the notion of user-centric design approaches does not always intuitively embed features that correspond to the users' characteristics and needs. A challenge met especially in current interactive systems is to dynamically adapt the content presentation and functionality of the system based on implicitly retrieved information about the user, aiming to improve usability and provide a positive user experience. Various research works exist in the literature that propose different approaches for Web adaptation and personalization, like [3] that proposes an approach for adapting user interfaces based on the cultural preferences of users, [4] that proposes an adaptive spellchecker and predictor for people with dyslexia that can adapt its model and interface according to the users' individual behavior, and [5] that proposes an implicit user modeling approach that automatically adapts the layout and position of virtual keyboards based on how and where users are grasping the tablet device.

Major commercial Web service providers nowadays have also shown an increased interest in providing personalized services to their users. These service providers have been offering personalized results and recommendations by employing various intelligent user modeling and adaptation techniques. Popular approaches for recommendation include collaborative filtering and content-based filtering [1, 6]. Collaborative filtering first collects and analyzes data about the users' interactions with the system or the users' preferences, and then predicts for the rest of the users their future preferences based on the similarity of their interests. Content-based filtering creates a user profile based on a weighted vector of the item features appearing in the content which is more frequently visited by the user. The weights indicate the importance of each feature to the user. Furthermore, various algorithms are employed to recommend new items that are similar to the weighted vector of the user. Various machine learning techniques are used to predict user preference or estimate the probability that users will like particular items, like cluster analysis, classification, decision trees, and artificial neural networks. Although the notion of personalization has found its way in users' everyday interactions in Web interactive systems, various research issues are still open with regards to the most influential factors of personalization, such as the behavioral drivers and navigation interaction of users in executing task-oriented reasoning processes. In addition, there is lack of understanding of the relation between individual styles and cognition levels and interactive behavior within interactive systems. Some illustrative examples are given below.

An interesting example is the case of users' interactions with online content, such as content included in encyclopedia articles. In that case, based on observations of human behavior and preference, the personalization process could influence both the way content is represented as well as the way the content is structured, and thus may have a significant impact on improving the users' experience. Assuming that the content of Web interactive systems can be presented in two ways, either a visual or a verbal representation illustrating the same content, and users may go through the content in a specific navigation pattern (or navigation behavior), this work suggests that individual differences in cognitive styles, which describe the way individuals perceive, process and organize information [7], might be applied effectively for facilitating the user modeling process of adaptive Web interactive systems [8]. The most widely accredited cognitive style dimensions are the Verbal/Imager dimension, that indicates the habitual approach and preference of users representing information verbally or graphically, and the Wholist/Analyst dimension, which describes the way individuals organize and process information in a holistic or an analytic approach [7, 11].

One of the main challenges in adaptive interactive systems is which characteristics of users should be included in the user model that is ultimately used for adaptation, and how to extract and represent these characteristics [9]. Especially in the case of modeling personality traits like cognitive styles, there are no easy ways to capture them during Web interactions. Prior work of the authors [10], has revealed that particular cognitive styles of individuals (i.e., Wholist/Analyst cognitive styles) can affect their navigation behavior in terms of linear/non-linear navigation behavior within Web-based environments based on specific navigation metrics that measure the degree of linearity an individual interacts with hyperlinks. In particular, various clustering techniques were performed on the Web navigation metrics obtained from user hyperlink interactions within a controlled Web environment. The clustering techniques used, aimed to group users that had similar interactions with hyperlinks, i.e., followed the same navigation in terms of linear and non-linear behavior, and further investigated whether there is a relationship with the cognitive styles of users, regarding the Wholist/Analyst dimension. The experiments were based on a user study of 106 individuals which navigated through the Web environment. The results revealed that the clustering process grouped consistently the users in the same groups based on their common navigation behavior. In addition, an intra-cluster analysis revealed that individuals that retain a global view of information (i.e., Wholists) had a more linear approach in navigation behavior (i.e., the users' interactions with hyperlinks tended to be sequential rather than scattered).

Furthermore, a number of research works exist in the related literature that aim to implicitly elicit cognitive styles of users based on their navigation behavior that focus primarily on the Wholist/Analyst dimension [31, 32, 33]. However, implicit user modeling approaches for eliciting the Verbal/Imager cognitive style of users are very scarce in the literature. In this respect, another challenge is to also investigate whether the Verbal/Imager cognitive style could be implicitly elicited based on the users' Web interaction data. Given that Verbal/Image cognitive style may be effectively correlated with content representation of hypermedia environments [8], a first approach for highlighting differences in cognitive styles based on users' Web interaction data would be to infer preference of users toward content representation based on the time they are active (implying interest) in verbal or visual representations of the same content. These correlations, once found, can improve dramatically the effectiveness of the personalized services and content delivery of Web systems.

In this context, main aim is to increase our understanding about the effect of cognitive styles of users on their navigation behavior, but as well as the content representation they prefer the most and increase their overall user experience. Thus, we extend our previous work as follows: i) we investigate whether there is an effect of the Verbal/Imager cognitive style dimension on users' preference of verbal or visual content representation, something which has not been reported in any prior work to the best of our knowledge, and ii) we further investigate the effect of the Wholist/Analyst cognitive style dimension on the navigation behavior of users based on a different representation scheme of Web navigation patterns than the one already proposed in [10] (i.e., based on sequence vectors). In particular, the more detailed analysis includes, apart from the investigation of the existence of a relation between linear/non-linear navigation behavior and the Wholist/Analyst dimension, whether the navigation path typically followed by the users of the same typology is similar or not. In other words, we investigate whether users with common cognitive characteristics have the tendency to follow exactly the same nodes in the hypermedia environment. Such a finding would further strengthen the range of valid metrics used until now for the implicit users' navigation behavior extraction in Web adaptive interactive systems.

The innovative aspects of this work lie in the introduction of a novel approach for implicitly capturing the users' interactions utilizing the structure of the Web environment, taking into consideration the distances between hyperlinks, the transition of users among hyperlinks utilizing sequence vectors, and the cognitive styles of users based on psychometric tests. The rest of this paper is organized as follows: in Section 2, we provide an overview of the related work and background theory. In Section 3 we present the user study conducted based on the proposed approach and we analytically discuss our results. Consequently, we conclude the paper in Section 4.

2. User Modeling

In this section we provide related material to user modeling and in particular, i) the underlying theory of cognitive styles utilized in this work, ii) an analysis of popular Artificial Intelligence techniques utilized by user modeling mechanisms in the context of adaptive interactive systems, and iii) related works on modeling cognitive styles utilizing data mining techniques on the users' Web interaction data.

2.1. *Cognitive Style Theory*

Research on cognitive styles is an area of human sciences to explain empirically observed differences in mental representation and processing of information. Different theories have been proposed over time suggesting that individuals have differences in the way they process and remember information. Due to the multi-dimensional nature of cognitive styles, a global definition has not been given to date. Nevertheless, in a global electronic survey of 94 individual style researchers and experts [11] from the European Learning

Styles Information Network (ELSIN) who were asked to comment on the state of the field and their own understanding of the phenomenon being studied, the majority agreed that *"cognitive styles are individual differences in processing that are integrally linked to a person's cognitive system. More specifically, they are a person's preferred way of processing (perceiving, organizing and analyzing) information using cognitive brain-based mechanisms and structures. They are partly fixed, relatively stable and possibly innate preferences"*.

The work of Riding and Cheema [12] is considered an important turning point for cognitive style research [13]. They conducted a survey of approximately thirty different cognitive styles and concluded that most of the proposed theories measured two broad style dimensions; i) a Verbal/Imager dimension that refers to how individuals process information and indicates their preference for representing information verbally (Verbals) or in mental pictures (Imagers), and ii) a Wholist/Analyst dimension that refers to how individuals organize information and indicates a preference of structuring information as a whole (Wholists) or structuring the information in segmented parts (Analysts). In addition, users with a Wholist cognitive style are supposed to take a linear approach in hypermedia navigation, whereas Analysts are supposed to take a non-linear approach in hypermedia navigation.

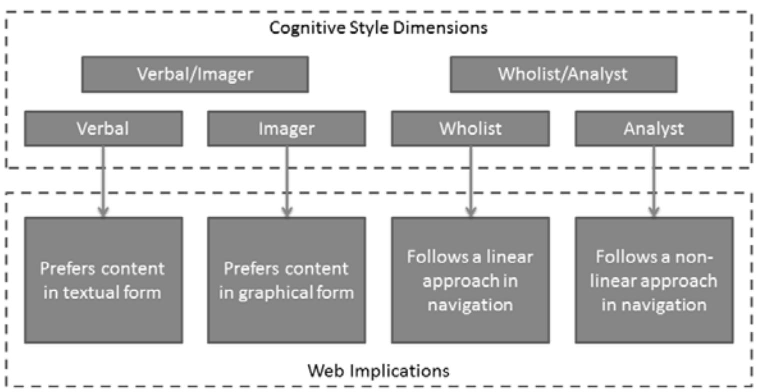

Figure 1. Riding's CSA Scale Mapping to Web Environments

Accordingly, Riding [7] proposed a computerized Cognitive Styles Analysis (CSA) test which highlights differences on these two broad dimensions. In this context, among the numerous proposed theories of cognitive styles [7, 14, 15], the proposed work utilizes Riding's CSA [7, 12] and the psychometric test since its implications may be mapped on Web environments as illustrated in Figure 1, and respond directly to different aspects of the Web information space [8]. In particular, the CSA implications may provide clear guidelines in the context of Web design, i.e., selecting to present visual or verbal content and structuring information flow in a wholistic or analytic manner.

2.2. *Artificial Intelligence Techniques for User Modeling*

User modeling is essential for adapting interactive systems which includes various characteristics about the users that are taken into consideration for the different adaptation effects provided. User modeling embraces various challenges, such as what user characteristics are important to be modeled and utilized by the adaptation mechanism, and how to extract these characteristics and further generate the user models. The simplest approach of user model generation is in the case where the information collected by the user is utilized as-is and remains unprocessed. For example, in an online video streaming system users might explicitly express their interest on specific movie genres which can be further used by simple rule-based mechanisms to adapt the interface by recommending movies that belong to the selected genres.

Given that user characteristics, needs and preferences might change over time, as well as, in many cases, users are unwilling to provide such information [16, 17, 18], explicit user model generation approaches usually result in user models becoming inaccurate over time. In this respect, a challenge is to implicitly and dynamically generate user models utilizing more sophisticated approaches, like cases where the browsing activities of users may be utilized by data mining and machine learning techniques to recognize regularities in user paths and integrate them in a user model. For example, in the context of the online video streaming system, mentioned above, the system would monitor the users' interaction data that might be useful for inferring information about the users, e.g., track how users rated movies of a particular genre, or how long they remained active in particular Web-pages.

A thorough literature review on how data mining techniques can be applied to user modeling in the context of Web personalization systems may be found in [19, 20, 21]. The data mining techniques reported enable pattern discovery through *clustering*, *classification*, *association rules*, and *Markov chains* for Web personalization purposes. *Clustering or fuzzy clustering techniques* group users together that share common characteristics or similar navigation behavior [22, 23, 24]. *Classification techniques* map user information (e.g., interaction data) into one of several predetermined classes which usually represent different user profiles [25]. *Association rule techniques* aim at generating associations and correlations among sets of items [26, 27]. *Markov chains* are used to represent the transitions of users within the Web environment [28, 29, 30] and they are introduced as a possible indication of which is the next page users might request to visit based on their current location and previous navigation paths. Thus, in the context of a Web application, representation schemes, like the ones in Markov chains can be utilized to represent the transition of users between Web-pages, using for example sequence vectors, and thus identify groups of users following same or similar paths.

2.3. *Data Mining for Eliciting Cognitive Styles of Users*

Various works have investigated the effect of cognitive styles on navigation behavior and learning patterns. Chen et al. [31] investigated how cognitive styles affect students' learning patterns in Web-based instruction programs utilizing statistical and data-mining techniques and consequently suggested design guidelines that take into consideration individual differences in cognitive styles for improving the learning process and user experience within Web-based instruction programs. Frias-Martinez et al. [32] utilized a number of clustering techniques to understand human behavior and perception in relation with cognitive styles, expertise and gender differences of digital library users. Hsu and Chen [33] investigated how learners' cognitive styles affect their navigation behavior through data mining techniques as well as analyzed how navigation behavior may influence performance in education environments.

The aforementioned works primarily focus on how individuals use search mechanisms and navigation tools (e.g., navigation maps, index of pages) and aim to cluster users based on the number of times each feature of the tools is used and further related to the users' cognitive styles regarding the Wholist/Analyst dimension. A challenging endeavor is to follow and monitor the interaction path of users during their experience with a Web environment. So, instead of monitoring usage, this paper proposes an alternative approach to user modeling by monitoring the users' sequence of links visited in a Web environment through an online tool that utilizes specific user interaction metrics aiming to examine how users navigate based on their cognitive styles regarding the Wholist/Analyst dimension. Additionally, in contrast to previous works that primarily focus only on the Wholist/Analyst dimension, this work also investigates the effect of cognitive styles regarding the Verbal/Imager dimension on users' content representation preference.

To this end, the overarching aim of this work is to increase our understanding and knowledge on supporting usable interaction designs with implicit user modeling based on users' cognitive styles and Web interaction data, through the use of a particular set of Artificial Intelligence techniques (i.e., clustering and sequence vector modeling).

3. User Study

This section explains the experimental procedure of the study, the process followed to obtain the cognitive styles of users and their Web interaction data, and the analysis and discussion of results. The analysis is based on a set of measures for cognitive styles, distance measures based on the structure of the Web environment and sequence vectors representing the transitions of users among hyperlinks.

3.1. *Participants*

A total of 78 undergraduate students of the University of Cyprus participated in the study (age of 17-25, with a mean age of 21 years old). The participants' native language was Greek, with knowledge of English as a second language. The participants first completed the cognitive styles elicitation process utilizing Riding's CSA test [7], and further navigated in a reproduced version of Wikipedia.org. With the aim to increase the users' navigation activity, participants were assigned 10 problem-based tasks whose answers could be found inside the Wikipedia articles to investigate their behavior in solving the problem-based tasks they had been assigned.

3.2. *Users' Cognitive Styles Elicitation*

A Web-based psychometric instrument, exploiting Riding's CSA [7], was developed that consists of two sub-tests. The first sub-test highlights individual differences in Wholist/Analyst cognitive style by requiring from the users to respond to 40 questions as true or false. In particular, 20 of the questions contain wholist-type stimuli that ask whether a pair of geometric shapes is identical or not (e.g., "Is shape X the same as shape Y?") (Figure 2), and the rest 20 questions contain analyst-type stimuli that ask whether a single geometric shape is part of another complex geometric figure ("Is shape X contained in shape Y?") (Figure 3). The response time is recorded for each user and a three step algorithm is applied to highlight the user's cognitive style as follows: i) calculate the average response time on each of the two sections (20 questions for the wholist-type stimuli, and 20 questions for the analyst-type stimuli), ii) calculate the ratio between the average response times on the wholist-type stimuli and analyst-type stimuli, and iii) associate the value of each subject's Wholist-Analyst ratio with a style category. A low ratio (≤1.02) classifies the participant as a "Wholist", a high ratio (>1.35) classifies the participant as an "Analyst", while a ratio in between the two end points classifies the participant as an "Intermediate" [7].

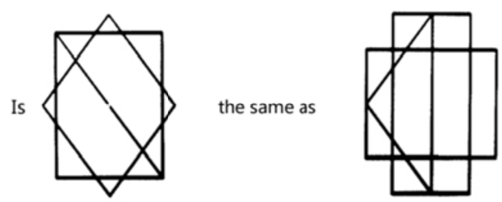

Figure 2. Example of a Wholist-type Stimuli

The second sub-test highlights individual differences in the Verbal/Imager cognitive style dimension by requiring from the users to respond to 48 statements, either being true or false. In particular, 24 statements require verbal reasoning by comparing two objects conceptually (e.g., "Are ski and cricket the same type?"). The rest 24 statements require visual reasoning by comparing the color of two objects (e.g., "Are cream and paper the same color?"). As in the first sub-test, the response time is recorded for each provided answer and a three step algorithm is applied to highlight the user's cognitive style as follows: i) calculate the average response time on each of the two sections (24 questions for the verbal-type stimuli, and 24 questions for the imager-type stimuli) of the CSA test, ii) calculate the ratio between the average response times on the verbal (conceptual category) and imagery (appearance) stimuli, and iii) associate the value of each subject's Verbal-Imager ratio with a style category. A low ratio (≤0.98) classifies the participant as a "Verbal", a high ratio (>1.09) classifies the participant as an "Imager", while a ratio in between the two end points classifies the participant as an "Intermediate" [7].

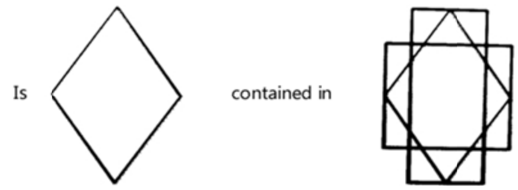

Figure 3. Example of an Analyst-type Stimuli

3.3. *Users' Web Interaction Monitoring*

A Web application based on Wikipedia.org was reproduced including additional functionality and content representations for the purposes of our study. In particular, the application monitored the Web interaction of users on the client-side utilizing a browser-based logging facility to collect the client-side usage data from the hosts accessing the Web application. The following Web interaction data were monitored: i) total view time of articles representing content either verbally or graphically, and ii) the hyperlink interactions. These are described next.

Monitoring Users' Content Representation Preference. Given that the Verbal/Imager dimension has implications on the content representation (verbal or visual) of Web environments, we monitored the users' preference towards content representation, by enriching the Web application to include both verbal-based content, i.e., content in textual form without images/visuals (Figure 4A), and image-based content, i.e., content represented with images/visuals and diagrammatical representations of text (Figure 4B). The participants had the option to either view the article in its textual version or in its graphical version. The total time spent in each version (viewing time) was recorded during the user's interaction with the system aiming to extract information about their preference towards a particular type of content representation.

Monitoring Users' Hyperlink Navigation Paths. Given that the Wholist/Analyst dimension refers to how individuals organize and perceive information, we assume that it might affect their approach to navigation. In this respect, we monitored the users' actual sequences with the hyperlinks of the Web application as well as calculated their linearity, i.e., whether a user navigated linearly from one link to the other or in a non-linear manner. Given that the structure of Wikipedia articles contains hyperlink anchors that point to specific sections within each article, we measured the actual sequence of visited hyperlinks and the linearity of user interactions with the hyperlinks within each article.

We have utilized two types of Web interaction metrics for representing the users' navigation paths with each article's hyperlinks; i) a Web navigation metric (proposed and used in [10]) that calculates the degree of linearity the users follow (linear or non-linear), and ii) sequence vectors to represent the actual transitions of users within each article. In this context, in order to represent the users' interactions, all hyperlinks within each article were automatically annotated with an attribute, meaningful to the system. In particular, a browser-based facility was developed that parsed a given HTML document and annotated each hyperlink with a unique incremental identifier, in the following format; *nav_n_m*, in which *n* identifies the article in which the user currently navigates and *m* the hyperlink clicked. Each time a user clicked on the annotated hyperlink, the unique identifier, as well as the time of hit was sent to the Web server. For example, for article with ID=1 consisting of 4 hyperlinks, the following identifiers were assigned to each hyperlink from top to bottom; nav_1_1, nav_1_2, nav_1_3, nav_1_4.

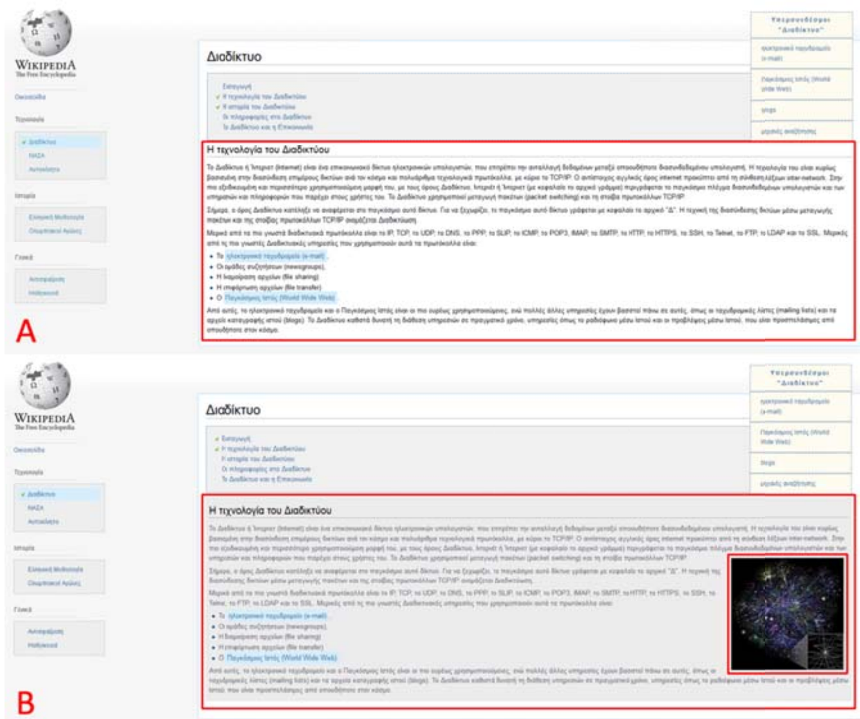

Figure 4. Verbal- and Image-based User Interface of the Web-site used in the Study

Next we describe how we utilized quantitative measures for representing each user's navigation behavior aiming to investigate what type of navigation metric (i.e., linearity navigation metric or navigation path) are more effective for implicitly eliciting the users' cognitive styles in respect to the Wholist/Analyst dimension.

Linearity Navigation Metric

We utilized the Absolute Distance of Links (ADL) metric [10], which is the total absolute distance between the links visited by a user, to measure the linearity of user interactions with the structured hyperlinks. In the equation below, which calculates the ADL metric,

x_i represents the identifier of links visited, i.e., i=1 is the first link visited (x_1 is equal to 1), i=2 the second (x_2 is equal to 2) and so on, and N is the number of total links clicked. Thus the distance between sequential links is assumed to be equal to 1.

$$ADL = \frac{|x_1 - 1| + \sum_{i=2}^{N} |x_i - x_{i-1}|}{N}$$

To better explain the metric used, we provide an example navigation, e.g., the click stream navigation pattern "nav_2_4 | nav_2_2 | nav_2_3", which indicates that the user visited article with ID=2 and then read the content of the fourth, second and third hyperlink of the navigation menu in the system. For this particular navigation, as defined above, the ADL metric is then calculated as: ADL=(|4-1|+|2-4|+|3-2|)/3=2. Accordingly, a high number of the metric indicates that the user followed a non-linear navigation behavior, whereas a small number of the metric indicates a linear navigation behavior.

Sequence Vector

Sequence vectors, inspired by Markov models are also used as metrics. Markov models or Markov chains are mathematical systems that consist of a discrete number of states and some known probabilities p_{ij}, where p_{ij} is the probability of moving from state i to state j. The representation power of Markov chains (particularly of sequence vectors) have been utilized to represent the navigation paths followed by the users through their interactions with hyperlinks in each article. The navigation sequence of a Web user is represented as a multidimensional probability matrix so that each element (i, j) in the sequence matrix indicates the proportion of visits to state j at the next transition, given the present state i. For example, given the sequence of visits of a user, $s1$ = 2-4-3-1-2, is represented by the sequence vector $v1$ = (0, 2, 4, 3, 1, 2, 0), where 0 indicates that the user starts the sequence and ends the sequence of navigation.

The sequence matrix is given as:

$$S1 = \begin{array}{cc} & \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} & & 1 & & \\ & & 1 & & \\ 1/2 & & & & 1/2 \\ & 1 & & & \\ & & & 1 & \end{pmatrix} \end{array}$$

Markov models have not been used to the extent of their probabilistic nature which could be beneficial for predicting, based on probabilities, the future navigation sequences of users.

3.4. *Design of Analysis*

Traditional statistics were first performed aiming to investigate whether there is an effect of cognitive styles on the content representation preference, and the navigation behavior of users. A series of analysis of variance (ANOVA) was performed in which the independent variable was the cognitive styles of users, and the dependent variables were the total time (in seconds) spent in each content representation version of the content (i.e., textual or graphical), and the Web interaction data metric (i.e., the ADL metric and the sequence vectors for each user).

Secondly, a series of cluster analyses was performed on all the interaction data obtained during the user study. The analyses included the following phases: First, we defined the optimum number of clusters using the two-step cluster analysis [34]. In the two-step process, in the first step, cases are assigned into pre-clusters and these pre-clusters are treated as single cases and in the second step using the hierarchical algorithm to cluster the pre-clusters. The analysis is based on an agglomerative hierarchical clustering method, which utilizes single-linkage clustering to determine which number of clusters is the optimal in each case. Particularly, we produced a range of 2-cluster solutions to 5-cluster solutions and then checked them one by one based on the clustering criterion of Schwarz's Bayesian Criterion (BIC) and selected the solutions that had the lowest BIC value, which represented the most well-separated clusters. After defining the number of optimal solutions (clusters) we utilized *k*-means clustering to obtain the cluster memberships, distance information, and the final cluster centers. In particular, the *k*-means clustering analysis was used to obtain the membership value ranging from 1 to the number of clusters and the distance from the cluster center for each user. The distance was measured using the Euclidean distance between each case and its classification center.

Since the data are derived from different users carrying out different navigations, they may be considered having the same probability distribution as the rest sequences of navigation, and thus all are mutually independent or generated independently. In addition, since the users were not directed in any way, the possible navigation patterns and user interactions with the user interface were close to a very large number. That is why *k*-means clustering was selected for the analysis, to avoid calculating all possible distances between all possible interactions.

3.5. *Analysis of Results*

In this section we present the results of the traditional statistics and clustering performed on the total view times indicating the content representation preference, the Web navigation metric and the sequence vector representations.

Users' Content Representation Preference and Cognitive Styles

The total time spent on the verbal version of the content and the total time spent on the graphical/diagrammatical version of the content (viewing time) was used to infer the users' preference toward a particular type of content presentation.

An ANOVA was performed to study the effect of cognitive styles (i.e., Verbal, Intermediate, Imager) on the total view time of each content representation type (i.e., textual and graphical content). A graphical illustration of the results is presented in Figure 5.

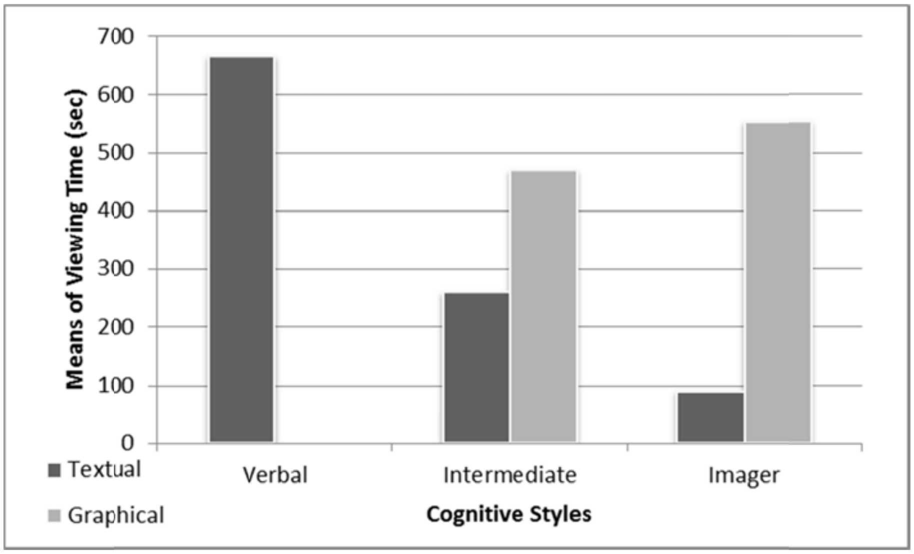

Figure 5. Content Representation View Time per Cognitive Styles Group

The ANOVA revealed a significant effect of cognitive styles of users on the view time of each content representation type ($F(2,151)=51.236$, $p<0.001$). Verbal users viewed significantly longer the textual content representation, whereas Imager users viewed significantly longer the graphical content representation. Such a result provides evidence that the particular cognitive styles investigated affect the users' preference toward the content representation, for these users, since each respective user group preferred the expected content representation based on their cognitive typology. In addition, the results obtained reveal that no significant differences were observed in the mean view time of both types of content representation regarding Intermediate users, which is in line with theory that Intermediates do not have a strong preference in information processing and toward specific content types.

Furthermore, Table 1 summarizes the preference of users (textual, visual or neutral form of the content) which is implicitly extracted based on their viewing time in each respective version of the content. A remarkable preference of textual content representation of Verbal users was found while both Intermediate and Imager users preferred the visual content.

Table 1. Users Cognitive Styles and Content Representation Preference

Preference / Cognitive Styles	Textual	Neutral	Visual
Verbal	37	0	0
Intermediate	3	2	15
Imager	1	1	17

A Pearson Chi-square test was also conducted to examine whether there is a relationship between the users' cognitive styles and preference toward a specific type of content. The results revealed that there is a significant relationship between these two variables (Chi square value=62.761, *df*=2, $p<0.001$). In particular, the users having Verbal cognitive style preferred the text-based version as they spent the majority of their session viewing time on the text-based version. Regarding users having Imager cognitive style, the majority preferred the visual/diagrammatical version. Finally, the majority of users having Intermediate cognitive style preferred the visual version. Given that Intermediate users do not have a strong preference toward a particular type of content and do not process efficiently either of the two content types (textual or graphical), a possible interpretation of this result might be related to the picture superiority effect in terms of attractiveness and delivery of information in a more efficient manner that might have affected their preference toward the visual/diagrammatical version compared to the text-based version.

Finally, we applied *k*-means clustering on the total viewing time for each content representation type of each user. Figure 6 illustrates the generated clusters and the distribution of users within each cluster based on their cognitive styles. Results reveal that all Verbal users were grouped in Cluster #1 in which they represent the majority. Most of the Imagers were assigned to Cluster #2, representing half the population of the cluster, whereas the rest of the population in Cluster #2 was Intermediate users. Looking more closely to the preference of these Intermediates, all of them spent more time in the visual/diagrammatical content representation (Mean: 584.93, Std. Dev.: 376.73).

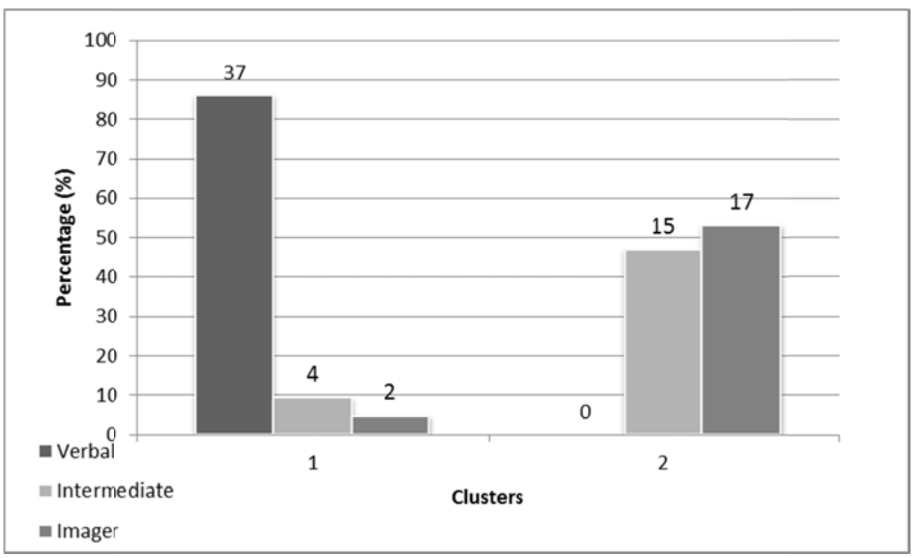

Figure 6. Clusters formed based on Total Viewing Time for each Cognitive Styles Group

Users' Navigation Behavior and Cognitive Styles

Next, we present the results obtained after the analyses performed on the content visit path followed by the users. In particular, both statistical and cluster analyses were performed on the Web navigation metric (ADL), whereas cluster analysis was performed on the sequence vectors.

Analysis of Web Navigation Metric Results

An ANOVA was performed to study the effect of cognitive styles on the navigation behavior of users based on the Web navigation metric (as defined in Section 3.3). Results revealed that the participants had significant differences in their navigation behavior based on their cognitive styles ($F(2,78)=14.004$, $p<0.001$). In particular, Wholists (Mean: 1.09, Std. Dev.: 0.55) and Intermediates (Mean: 1.12, Std. Dev.: 0.5) followed a more linear navigation since the majority of them had a Web navigation metric very close to 1. Whereas the majority of Analysts (Mean: 2.02, Std. Dev.: 0.98) had a Web navigation metric greater than 2, indicating a non-linear navigation behavior. The existence of high standard deviation values of the metric indicates that the relation found between navigation behavior and cognitive styles does not hold for all users. This is however expected, considering the complex nature of human-related data such as the users' cognitive styles and navigation behavior, for which it is extremely hard to obtain general relations holding in every single case. In this respect, further studies need to be conducted to reach to more concrete conclusions about the relations between the Wholist/Analyst dimension and navigation behavior.

K-means clustering was also performed on the Web navigation metric value of each user to investigate the feasibility of eliciting the cognitive styles of users based on their navigation behavior in Web environments. Figure 7 illustrates the generated clusters and the distribution of users within each cluster according to their cognitive styles. Results reveal that the clustering performed grouped the users in different clusters, however, with varying cognitive styles. In particular, Cluster #1 contains in the majority Intermediates, and half of the Wholists and Analysts of the total sample. Cluster #2 includes the rest of the Analysts and Wholists and a few Intermediates. In this respect, no safe conclusions can be drawn whether users with similar cognitive styles have the same navigation behavior. Nevertheless, taking a closer look to the Intermediates' cognitive style ratios in Cluster #1, we observed that these users tended to be Wholists (i.e., their ratio were quite low and thus may be treated as Light Wholists) indicating that the majority of these users grouped in Cluster #1 (i.e., the Wholists and Light Wholists) had similar linear navigation behavior. Cluster #2 contains users of variant cognitive typologies since that particular cluster contains users that did not have any clear/extreme navigation behavior in terms of linear or non-linear behavior, and therefore we cannot infer anything about the existence of a relation between their cognitive styles and navigation behavior.

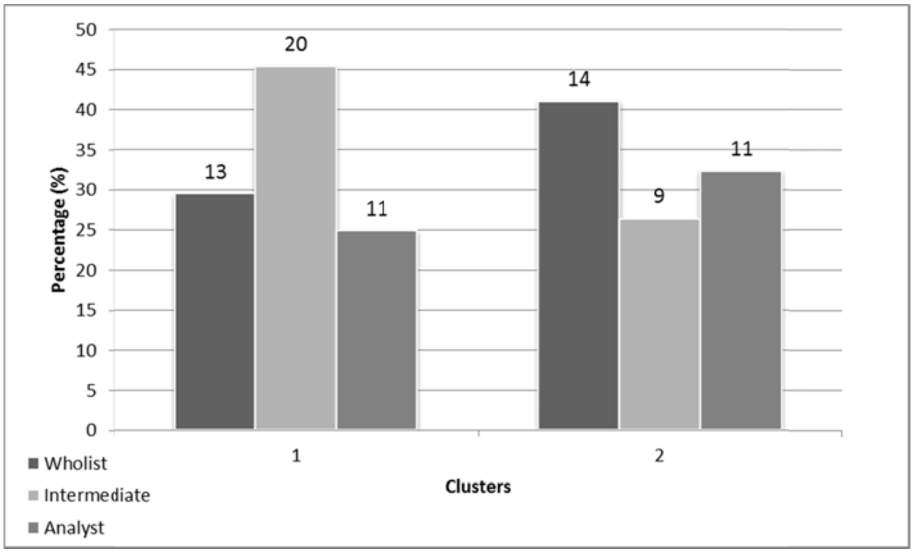

Figure 7. Clusters formed based on Web Navigation Metric for each Cognitive Styles Group

Analysis of Sequence Vectors Results

An analysis of the sequence matrices was performed to investigate whether the representation of the users' navigation behavior with sequence vectors could be effectively used for clustering users performing similar transitions. The analysis involved examination of the typologies of users in each cluster based on Riding's CSA. In Figure 8 we visualize the clusters produced based on the sequence vectors and the number of users of each cognitive style group.

Accordingly, the clustering method distinguished users in two clusters indicating that users had differing navigation behavior. An intra-cluster analysis based on the users' cognitive styles revealed that all clusters include users with varying cognitive styles indicating that users with similar cognitive styles do not follow always an identical or at some high degree, a similar navigation path. This might be justified by the fact that the actual navigation path might be influenced by other factors (i.e., motivation, interests) than cognitive styles which primarily affect at a more high level the general approach towards organizing information and navigating in a hypermedia environment. Nevertheless, a closer look at the results provides some indications about the similar navigation behavior of users grouped in the clusters, since in Cluster #1, the majority of users are Wholists and Intermediates compared to the number of Analysts, and in Cluster #2, Analysts and Intermediates outnumber Wholists. Finally, some superiority seems to exist when it comes to clustering using sequence vectors over the Web navigation metric that measures linear/non-linear navigation behavior. However, this needs to be confirmed with further user studies.

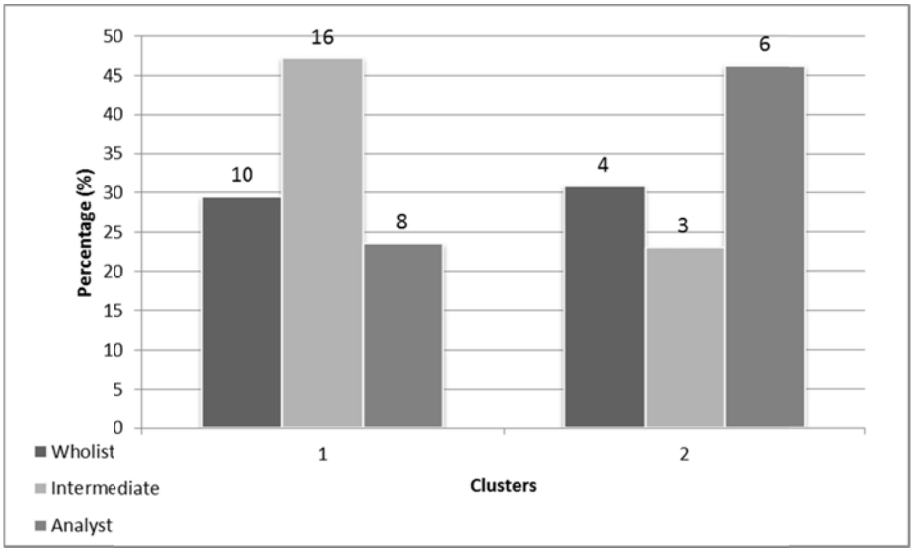

Figure 8. Clusters formed based on the Sequence Vectors for each Cognitive Styles Group

Summarizing, the analysis revealed that users of a particular cognitive style have a clear preference toward a particular content version indicating a relation between cognitive styles and the viewing time (preference) of different types of content representation (textual vs. graphical). Furthermore, the analyses of navigation behavior indicated that differences exist in navigation style between users, and some conclusions may be drawn about the relationship between users' navigation and cognitive styles. In particular, comparing the clustering techniques we applied on the Web navigation metric (ADL) and on a new representation method of the navigation behavior (using sequence vectors), we have obtained improved groupings of users in meaningful clusters, even though this occurred for half of the users in the study. In this respect, further investigations are needed to reach to more concrete relationships than the ones found.

4. Conclusions

Processing of human-related data, and particularly implicit data captured during the interaction of users with Web systems, is particularly intriguing since the human aspect is complicated by nature and hardly understood. Accordingly, the overarching aim of this work was to increase our understanding on mining human-related behavior through navigation interaction of users with Web environments based on Artificial Intelligence techniques. Accordingly, users' Web interaction data were utilized and typical statistical methods and clustering techniques were performed, with the aim to identify patterns, trends, similarities in navigation behavior as well as user preference toward different types of content representation, and investigate the relation, if it exists, to their cognitive styles. Such a finding could provide a promising direction toward the identification of adaptation rules, based on the abovementioned relationship, for designing better, implicit, automatic and dynamic user-centric Web interactive systems.

In sum, results revealed that cognitive styles of users (taking into account the Verbal/Imager dimension) could be inferred, i.e., a strong significant relationship was found between the users' cognitive styles and preference toward a specific type of content, based on their total viewing time of Web-pages that contain visual or verbal information. A practical implication of this finding, is that user modeling techniques could be improved so that they would implicitly create the user model, by tracking the users' total viewing time on a particular Web-page, that would be priory categorized based on its content (text- or image-based) and further utilized to infer the users' preference toward a particular type of content representation, and their cognitive style based on the Verbal/Imager dimension.

Furthermore, regarding the analysis of navigation behavior and cognitive styles, using the Web navigation metric, users were found to have significant differences in their navigation behavior based on their cognitive style with respect to the Wholist/Analyst dimension. However, the clustering performed grouped the users in different clusters, but, with varying cognitive styles, so further user studies need to be conducted before any safe conclusions may be drawn. Nevertheless, the clustering of users, based on the sequence vectors has shown some promising results and was effective to locate users with similar navigation behavior and in the same cognitive typology. This observation was made since consistently and in many cases the clustering technique grouped in the same clusters homogeneous users based on their navigation behavior and cognitive styles.

Another practical implication of this work could be the creation of an improved personalization engine that would implicitly and dynamically identify the cognitive styles of Web-site visitors based on their navigation behavior and view time of particular pages, and further feed an adaptation engine with the user models providing different adaptation effects. Content adaptation effects based on different cognitive styles of users could for example present content in diagrammatical representation in case of an Imager user, or present content in a Verbal representation in case of a Verbal user. Adaptive navigation support could also be provided to users with particular cognitive styles by adapting the sequence of hyperlinks to support a holistic and guided navigation approach for Wholists, or a more analytic and scattered navigation approach for Analysts.

Given that this work is an initial and indeed challenging approach to understand the relation of cognitive styles and users' Web interaction data, further studies need to be conducted in order to reach to more concrete conclusions about the effect of cognitive styles of users on their navigation behavior and content representation preference.

Our future research steps include to further investigate the needs, preferences and behaviors of users by analyzing their interactions in other Web environments and in particular domains of discourse, such as educational, commercial and collaborative. A particular challenge in these environments is the analysis of their more complex structures, and also take into consideration effects caused by the social networks, collaborative-filtering, advertisements and other factors that cause implications. Another interesting issue includes additional analysis of visual objects such as, drop down lists, search engines and drag and drop features that could influence the users' interactions. Finally, future work can include in the tool proposed, Markov models theory or Bayesian networks for the prediction of the behavior of new users in the system, without monitoring their interaction from beginning to end, and implicitly creating their user model using advanced Artificial Intelligence methods.

Acknowledgements. The work is co-funded by the PersonaWeb project under the Cyprus Research Promotion Foundation (ΤΠΕ/ΠΛΗΡΟ/0311(ΒΙΕ)/10), and the EU projects Co-LIVING (60-61700-98-009) and SocialRobot (285870). The work was also partially supported and carried out during the tenure of an ERCIM "Alain Bensoussan" Fellowship Programme. The research leading to these results has received funding from the *European Union* Seventh Framework Programme (*FP7/2007-2013*) under *grant agreement* no. 246016.

References

1. P. Brusilovsky, A. Kobsa, and W. Nejdl, *The Adaptive Web: Methods and Strategies of Web Personalization* (Springer-Verlag, Heidelberg 2007)
2. P. Brusilovsky, Adaptive Hypermedia. *J. User Modeling and User-Adapted Interaction*, 11(1,2) (2001) 87-110.
3. K. Reinecke, and A. Bernstein, Improving Performance, Perceived Usability, and Aesthetics with Culturally Adaptive User Interfaces. *J. Transactions of Computer-Human Interaction*, 18(2), (2011) Article 8, 29 pages.
4. A. Li, L. Sbattella, and R. Tedesco, PoliSpell: An Adaptive Spellchecker and Predictor for People with Dyslexia, in *Proc. of Int. Conference on User Modeling, Adaptation and Personalization*, (Springer-Verlag, 2013), pp 302-309.
5. L. Cheng, H. Liang, C. Wu, M. Chen, iGrasp: Grasp-based Adaptive Keyboard for Mobile Devices, in *Proc. of Int. Conference on Human Factors in Computing Systems*, (ACM press, 2013), pp. 3037-3046.
6. J. Konstan, and J. Riedl, Recommender Systems: From Algorithms to User Experience, *J. User Modeling and User-Adapted Interaction*, 22(1-2) (2012) 101-123.
7. R. Riding, *Cognitive Style Analysis - Research Administration*, Learning and Training Technology (Birmingham, UK, 1991)
8. P. Germanakos, N. Tsianos, Z. Lekkas, C. Mourlas, and G. Samaras, Realizing Comprehensive User Profile as the Core Element of Adaptive and Personalized Communication Environments and Systems, *The Computer Journal*, 52(7) (2008) 749-770.
9. P. Brusilovsky, and E. Millán, User Models for Adaptive Hypermedia and Adaptive Educational Systems, in *The Adaptive Web*, eds. P. Brusilovsky, A. Kobsa, and W. Nejdl (Springer-Verlag, Heidelberg, 2007), pp. 3-53.
10. M. Belk, E. Papatheocharous, P. Germanakos, G. Samaras, Modeling Users on the World Wide Web based on Cognitive Factors, Navigation behavior and Clustering Techniques. *J. of Systems and Software*, 86 (12), (2013) 2995-3012.
11. E. Peterson, S. Rayner, and S. Armstrong, Researching the Psychology of Cognitive Style and Learning Style: Is There Really a Future?. *J. Learning and Individual Differences*, 19(4) (2009) 518-523.
12. R. Riding, and I. Cheema, Cognitive styles - An Overview and Integration, *J. Educational Psychology*, 11(3/4) (1991) 193-215.

13. E. Peterson, I. Deary, and E. Austin, A New Reliable Measure of Verbal-Imagery Cognitive Style, *J. Personality and Individual Differences*, 38 (2005) 1269-1281.
14. R. Felder, and L. Silverman, Learning and Teaching Styles in Engineering Education, *Engineering Education*, 78 (1988) 674-681.
15. H. Witkin, C. Moore, D. Goodenough, and P. Cox, Field-dependent and field-independent cognitive styles and their educational implications, *Review of Educational Research*, 47(1), (1977) 1-64.
16. S. Gauch, M. Speretta, and A. Chandramouli, A. Micarelli, User Profiles for Personalized Information Access, in *The Adaptive Web*, eds. P. Brusilovsky, A. Kobsa, and W. Nejdl (Springer-Verlag, 2007), pp. 54-89.
17. G. Jawaheer, M. Szomszor, and P. Kostkova, Comparison of Implicit and Explicit Feedback from an Online Music Recommendation Service, in *Proc. of Int. Workshop on Information Heterogeneity and Fusion in Recommender Systems*, (ACM press, 2010), pp. 47-51.
18. A. Wærn, User Involvement in Automatic Filtering: An Experimental Study, *J. User Modeling and User-Adaptive Interaction* 14(2-3) (2004) 201-237
19. M. Eirinaki, and M. Vazirgiannis, Web Mining for Web Personalization, *J. ACM Transactions on Internet Technology*, 3(1) (2003) 1-27.
20. D. Pierrakos, G. Paliouras, C. Papatheodorou, and C. Spyropoulos, Web Usage Mining as a Tool for Personalization: A Survey, *J. User Modeling and User-Adapted Interaction*, 13(4) (2003) 311-372.
21. B. Mobasher, Data Mining for Web Personalization, in *The Adaptive Web*, eds. P. Brusilovsky, A. Kobsa, and W. Nejdl (Springer-Verlag, Heidelberg, 2007), p. 90-135.
22. O. Nasraoui, M. Soliman, E. Saka, A. Badia, and R. Germain, A Web Usage Mining Framework for Mining Evolving User Profiles in Dynamic Web Sites, *J. IEEE Transactions on Knowledge and Data Engineering*, 20(2) (2008) 202-215.
23. G. Castellano, and M. Torsello, Categorization of Web Users by Fuzzy Clustering, in *Proc. Int. Conf. Knowledge-Based Intelligent Information and Engineering Systems*, (Springer-Verlag, Heidelberg, 2008), pp. 222-229.
24. G. Castellano, A. Fanelli, C. Mencar, and M. Torsello, Similarity-Based Fuzzy Clustering for User Profiling, in *Proc. of Int. Conf. Web Intelligence and Intelligent Agent Technology Workshop*, (IEEE/WIC/ACM, 2007), pp. 75-78.
25. X. Wu, V. Kumar, J.R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. McLachlan, A. Ng, B. Liu, P. Yu, Z. Zhou, M. Steinbach, D. Hand, and D. Steinberg, Top 10 Algorithms in Data Mining, *J. Knowledge Information Systems,* 14(1) (2007) 1-37.
26. G. Linden, B., Smith, and J. York, Amazon.com Recommendations: Item-to-Item Collaborative Filtering, *J. IEEE Internet Computing*, 7(1) (2003) 76-80.
27. X. Su, and T. Khoshgoftaar, A survey of collaborative filtering techniques, *Advances in Artificial Intelligence*, Article 4 (2009) 19 pages.
28. S. Parka, N. Sureshb, and B. Jeonga, Sequence-based Clustering for Web Usage Mining: A New Experimental Framework and ANN-enhanced K-means Algorithm, *Data & Knowledge Engineering*, 65(3) (2008), 512–543.
29. I. Cadez, D. Heckerman, C. Meek, P. Smyth, and S. White, Visualization of Navigation Patterns on a Web site using Model-based Clustering, in: *Proc. of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (ACM press, 2000) 280–284.
30. Q. Yang, J.Z. Huang, M. Ng, A Data Cube Model for Prediction-based Web Prefetching, *J. Intelligent Information Systems,* 20(1) (2003) 11–30.
31. S. Chen, and X. Liu., An Integrated Approach for Modeling Learning Patterns of Students in Web-Based Instruction: A Cognitive Style Perspective, *J. ACM Transactions on Computer-Human Interaction* 15(1) (2008) Article 1, 28 pages.

32. E. Frias-Martinez, S. Chen, R. Macredie, and X. Liu, The Role of Human Factors in Stereotyping Behavior and Perception of Digital Library Users: A Robust Clustering Approach, *J. User Modeling and User-Adapted Interaction*, 17(3) (2007) 305-337.
33. Y. Hsu, and S. Chen, Associating Learners' Cognitive Style with their Navigation Behaviors: a Data-mining Approach, in *Proc. Int. Conf. Human-Computer Interaction: Users and Applications*, (Springer-Verlag, 2011), pp. 27-34.
34. M. Norusis, SPSS 13.0 Statistical Procedures Companion. Upper Saddle-River, N.J.: Prentice Hall, Inc. (2004)
35. P. Germanakos, E. Papatheocharous, M. Belk, and G. Samaras, Data-driven User Profiling to Support Web Adaptation through Cognitive Styles and Navigation Behavior, in Proc. of Mining Humanistic Data Workshop, (Springer-Verlag, Heidelberg, 2012), pp. 500-509.