The Journal of Systems and Software 86 (2013) 2995–3012

Contents lists available at ScienceDirect

The Journal of Systems and Software

journal homepage: www.elsevier.com/locate/jss

Modeling users on the World Wide Web based on cognitive factors, navigation behavior and clustering techniques

Marios Belk*, Efi Papatheocharous, Panagiotis Germanakos, George Samaras

Department of Computer Science, University of Cyprus, Nicosia, Cyprus

ARTICLE INFO

Article history:
Received 14 April 2012
Received in revised form 8 March 2013
Accepted 15 April 2013
Available online 7 May 2013

Keywords:
User modeling
Cognitive styles
Web navigation behavior

ABSTRACT

This paper focuses on modeling users' cognitive styles based on a set of Web usage mining techniques on user navigation patterns and clickstream data. Main aim is to investigate whether specific clustering techniques can group users of particular cognitive style using measures obtained from psychometric tests and content navigation behavior. Three navigation metrics are proposed and utilized to find identifiable groups of users that have similar navigation patterns in relation to their cognitive style. The proposed work has been evaluated with two user studies which entail a psychometric-based survey for extracting the users' cognitive styles, combined with a real usage scenario of users navigating in a controlled Web 2.0 environment. A total of 106 participants of age between 17 and 25 participated in the study providing interesting insights with respect to cognitive styles and navigation behavior of users. Studies like the reported one can be useful for modeling users and assist adaptive Web 2.0 environments to organize and present information and functionalities in an adaptive format to diverse user groups.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

During the last decade, Web 2.0 has ingrained itself into everyday life and has contributed to the exponential increase of internet usage since the early 2000s. Its fundamental concept as a medium for collaboration and sharing of information has generated extensive enthusiasm driving many of the world's markets. Within this realm, adaptivity considerations in modern Web-based interactive systems are considered of paramount importance as it is known that traditional static interactive Web-based systems treat all users the same way, being unable to satisfy the heterogeneous needs and preferences of users (Brusilovsky, 2001). This often results in users experiencing orientation difficulties in the information space or being overwhelmed by redundant information. To elaborate on this by giving examples, a static electronic encyclopedia usually offers the same functionality and content to readers with different knowledge and interests. Similarly, a traditional e-commerce Web-site offers the same product catalog and tools to customers with different needs and preferences. Finally, a traditional educational hypermedia system presents the same static content to students with widely differing educational goals and educational background.

In this respect, adapting the functionality and content, of any Web-based interactive system, to satisfy the users' needs and increase their level of understandability and acceptability in an intuitive manner and empower them to complete specific tasks more efficiently and effectively is a challenging endeavor. It entails understanding and modeling human behavior for diverse user groups, with regard to structural and functional user requirements, which needs to be translated into usable computer–human interaction designs and workflows, whilst minimizing the overall users' cognitive, perceptual and learning load. Adaptive interactive systems (Brusilovsky et al., 2007; Brusilovsky, 2001; Bra et al., 2000; Schneider-Hufschmidt et al., 1993) provide an alternative to the "one size fits all" approach of static user interfaces by adapting the interactive system's structure, terminology, functionality and presentation of content to users' perceptions, needs and preferences, aiming to increase the usability of the interface and provide an improved user experience.

One of the distinctive features of an adaptive interactive system is its user model. The user model is a representation of static and dynamic information about an individual that is utilized by the adaptive interactive system aiming to provide adaptation effects (i.e., the same system can look different to users with different user models) (Brusilovsky and Millán, 2007; Frias-Martinez et al., 2005). To better explain this through some examples, an information retrieval system may adaptively select and prioritize the most relevant items to the user's goals and/or interests. An educational hypermedia system may provide adaptive navigation support by manipulating the links based on the user's knowledge and

* Corresponding author. Tel.: +357 22 892675.
E-mail addresses: belk@cs.ucy.ac.cy (M. Belk), efi.papatheocharous@cs.ucy.ac.cy (E. Papatheocharous), pgerman@cs.ucy.ac.cy (P. Germanakos), cssamara@cs.ucy.ac.cy (G. Samaras).

0164-1212/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jss.2013.04.029

learning goals. A security and privacy-preserving related mechanism in a commercial Web system may present the content adaptively to the user's level of knowledge toward security terms (e.g., provide novice users with personalized security information awareness by using simplified security terms and additional explanations).

The mechanism used for user modeling can be based on explicit or implicit information gathering approaches. Explicit information is provided directly by the user, usually through Web registration forms, questionnaires, or specially designed psychometric instruments. On the other hand, implicit information is extracted by the system automatically to infer characteristics about the user and is usually obtained by tracking the user's navigation behavior by interacting with the system. For example, such implicit information can be extracted based on the time spent on a particular Web-page by a user, which can be used to infer the interest of the user toward the main subject of that Web-page. Various research works have attempted to investigate the most effective source of information for user modeling (Gauch et al., 2007; Jawaheer et al., 2010; Wærn, 2004). Based on Gauch et al. (2007) it is yet not clear-cut whether implicitly created models are more or less accurate than explicitly created models. Nevertheless, since implicit information gathering does not affect the human computer interaction or the users' cognitive load (Gauch et al., 2007), it seems to be the preferable approach for collecting information about users. On the other hand, this approach is much more complex than explicit user feedback since in most cases the data obtained may be imprecise, incomplete and/or heterogeneous.

An increasing interest, apart from research-oriented systems, has been observed during the last few years from major commercial Web 2.0 service providers such as Google (http://www.google.com), Bing (http://www.bing.com), and Amazon (http://www.amazon.com) that provide personalized results and recommendations, by employing various user modeling and adaptation techniques. In this respect, the notion of personalization has finally found its way in users' everyday interactions in Web interactive systems. Nevertheless, there is a need for more research on measuring the actual benefit for users, instead of merely developing elaborated user modeling and adaptation techniques. Even if personalization is the key to more efficient interactions and enhancing user experience, there is one undeniable issue to be resolved: how and why could users benefit even more? Individuals are certainly different from each other, but which would be the underlying theories that could guide research endeavors in producing measurable gains? A first approach would be to identify the levels in which individuals demonstrate considerable differences, such as demographics, social, cognitive and mental abilities, personality, goals, needs, and experiences, and to build a cohesive user model by including characteristics that could be important in adapting Web 2.0 interactive systems, with the aim to improve the user experience. Since content and functionality of Web 2.0 interactive systems is either presented visually or verbally, and users may have specific navigation behavior, e.g., holistic or analytic approach of navigation, this work suggests that cognitive styles of users, which describe the way individuals, perceive and organize information, might be applied effectively on designing adaptive Web 2.0 interactive systems. In this context, a Web environment's information presentation and structure adapted to users' cognitive style may increase the usability of a system in terms of efficiency and effectiveness, as well as provide a positive user experience (Germanakos et al., 2008).

To this end, the work presented is primarily driven by the need to apply User-Centered Design (UCD) methodologies related to the design and evaluation of adaptive mechanisms, and contributes toward this direction by proposing a user modeling approach for highlighting cognitive styles of users with explicit and implicit information gathering approaches by utilizing data analysis techniques. In particular, a specific psychometric measurement is used to highlight differences in cognitive styles of users, combined with data analysis techniques applied on the users' navigation patterns. Main objectives of the paper are to: (i) study the relation between users' cognitive styles and navigation behavior, (ii) investigate whether specific data analysis techniques can group users of particular cognitive style using measures obtained from psychometric tests, and (iii) propose navigation content metrics to find identifiable groups of users that have similar navigation patterns. The identification of users with specific cognitive and navigation style will ultimately help in defining various adaptation mechanisms which are required to be assembled to target a different user interface experience in Web-based environments for various cognitive typologies of users.

Overarching aim is to drive this research toward the design and development of a comprehensive adaptive interactive system which will be conceptually composed of two interconnected components; the user modeling and the adaptation component. The user modeling component will entail data about its users and their interactions. This information will be provided to the system either explicitly by the users (e.g., through registration forms or psychometric tests, etc.) and/or implicitly retrieved through user's interactions with the system aiming to enrich the user model and to infer information which is considered valuable in order to provide adaptive features (e.g., track the users' navigation behavior, and further infer their cognitive style). In this context, various data and cluster analysis techniques will be performed on the raw data acquired in order to generate the actual user models which, combined with various decision making and adaptation mechanisms, decide on the adaptation effects to be performed that will be further communicated to the adaptive user interface (e.g., in case users have a holistic approach in information organization and follow a linear approach in navigation, then provide personalized navigation tools that will assist the users process and organize information more efficiently and effectively).

The paper is organized as follows. In Section 2, we provide an overview of user modeling mechanisms giving emphasis on explicit and implicit gathering approaches, and cognitive-based user modeling systems. In Section 3 we present a user study based on the proposed approach. Consequently, we conclude the paper and describe our directions of future work in Section 4.

2. User modeling mechanisms

The ability of adaptation in interactive systems heavily depends on successful user modeling. A user model is created through a user modeling mechanism in which unobservable information about a user is inferred from observable information from that user (Frias-Martinez et al., 2005), for example, using the interactions with the system (i.e., time being active on a Web-page, buying history, ratings of products, bookmarked or saved content, etc.). User models can be created utilizing explicit information from the user, i.e., user guided modeling, and/or implicit information, i.e., dynamic user modeling (Gauch et al., 2007; Frasincar et al., 2004; Houben et al., 2004). These two categories are discussed in the next sections.

2.1. User guided modeling

User guided modeling methodologies rely on personal information provided by the users, typically via registration forms. The data collected usually contain demographic information (i.e., age, gender, and profession), interests and/or preferences. Common techniques for obtaining explicit information that allows specification of the users' model include the use of checkboxes, drop-down

lists, or text fields where users express freely their opinion. All these techniques have the advantage that the format of the replies are standardized but the main drawback is that the user is aware that the system is storing this information and usually the process may be disrupted due to unwillingness of the user to provide the information, lack of trust or time to participate in the process. Also, the results from these techniques are human-error prone since if the questions are not carefully designed then they might be inaccurate, inconclusive, or at worst, deceptive.

User guided modeling approaches are commonly utilized for customizing user interfaces. In this case, a collection of user preferences are used to create a user profile and the services provided adapt in order to increase information accessibility. For instance, the iGoogle application (iGoogle, 2012), explicitly asks the users to provide their personal information which is stored to create user profiles. The Web-site content is then dynamically organized based on the users' preferences.

An important drawback of customization approaches is that users may not accurately or fully report their preferences and characteristics. Furthermore, most interactive systems utilizing such approaches never invoke the user to update the information and rarely have intelligent mechanisms behind to identify that something has changed in the users' preferences. This usually results in profiles remaining static even though the user's interests may change over time. As a consequence, the profile may become highly inaccurate over time.

2.2. Dynamic user modeling

User models could be also dynamically generated based on implicit information, such as the navigation behavior of users. These mechanisms are transparent to the user and do not add disruption or require any additional effort by the user during the process of interacting with the system for constructing the models. Kelly and Teevan (2003) provide an overview of the most popular mechanisms for dynamically collecting implicit user information. Gauch et al. (2007) also summarize different approaches to implicit user information collection.

The most common source of information about users is their browsing history from which the users' interests are extracted. Browsing history of users contains URLs visited by the user and the date/time of the visits. Accordingly, meaningful information could be extracted based on this information, e.g., number of visits to a particular URL and the time spent in that Web-page. Browsing histories could be collected in two ways: users sharing their browsing caches on a periodic basis (Gauch et al., 2007), or users installing a proxy server that acts as their gateway to the Internet, thereby capturing all Web traffic generated by the user (Trajkova and Gauch, 2004). The first technique utilizes the Web browser's cache system that stores the user's Web browsing history. This technique does not require any installation of specific software. However, in order to extract meaningful information from the collected data, the user is required to upload the cache periodically. In the second technique, proxy servers allow easily capturing information without placing any major burden on the user because they only require an initial setup and do not require any software to be maintained or updated afterwards on the user's desktop. An important drawback however is that no user model can be created without having a specific proxy enabled by the user.

Another approach to collect implicitly information while the user navigates in an interactive system is through the usage of agents (e.g., browser plugins). Browser agents can be installed on the user's desktop computer and are able to capture all of the activities the user performs while browsing. Apart from collecting the user's browsing history (i.e., URLs visited), browser agents accurately collect information about the actions performed on a Web-page, such as bookmarking and downloading to disk. Accordingly, based on this additional information about the user's browsing activity, agents may suggest links on the current page that might be of interest. An important drawback of this approach is that it requires specialized software to be installed.

Enhancements of browser agents are desktop agents which are commercial toolbars that include personalized features with the aim to help users organize their browsing activity stored in their desktop caches. Furthermore, navigation activity for desktop agents is not limited to the Web, but also includes access of users on their local computer, e.g., personal folders and documents. Such search tools are implemented in applications like Google Desktop Search (2012) and Stuff I've Seen (Dumais et al., 2003).

2.3. Generating user models

The simplest approach of user model generation is in the case where the information collected by the user is used as-is and remains unprocessed. For example, users might explicitly express their interest on specific topics of a news publishing system which will be further used by simple rule-based mechanisms to adapt the interface by displaying the selected topics on the top of the users' interface. More intelligent approaches for generating user models include 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. A thorough literature review on how data mining techniques can be applied to user modeling in the context of personalization systems can be found in (Eirinaki and Vazirgiannis, 2003; Pierrakos et al., 2003). The data mining techniques mentioned enable pattern discovery through clustering and classification, association rules (or association discovery) and sequence mining (or sequential pattern discovery). They represent popular approaches appearing in the data mining literature. In addition, Mobasher (2007) describes data mining algorithms based on clustering, association rule discovery, sequential pattern mining, Markov models and probabilistic mixture and hidden (latent) variable models for Web personalization purposes.

Nowadays, the process of Web user modeling has become attached to automated data mining or knowledge discovery techniques due to the large volumes of available user data on the Web (Nasraoui et al., 2008). Nasraoui et al. (2008) perform clustering on user sessions to place users in homogeneous groups based on the similar activities performed and then extract specific user profiles from each cluster. Clustering techniques are also used in order to divide users into segments containing users with similar navigation behavior. Using a similarity metric, a clustering algorithm groups the most similar users together to form clusters. Because optimal clustering over large data sets is impractical, most applications use various forms of greedy cluster generation. These algorithms typically start with an initial set of segments, which often contain one randomly selected user. Then, they repeatedly match users to the existing segments. Once the algorithm generates the segments, it computes the users' similarity to vectors that summarize each segment, chooses the segment with the strongest similarity and classifies the user accordingly. Some algorithms classify users into multiple segments and describe the strength of each relationship (Perkowitz and Etzioni, 2000). The same concept is found within fuzzy clustering techniques, examples of which include the work of Castellano and Torsello (2008) that categorized users based on the evaluation of similarity between fuzzy sets using a relational fuzzy clustering algorithm and Castellano et al. (2007) that derived user profiles by analyzing user interests. Variations of fuzzy clustering methods include Fuzzy c-medoids, Fuzzy c-trimmed-medoids, Relational Fuzzy Clustering-Maximal Density Estimator (RFC-MDE)

algorithm, hierarchical clustering approaches, which were applied to group user sessions (Joshi and Joshi, 2000; Fu et al., 1999).

To this end, the abovementioned works primarily focus on applying data mining and machine learning techniques for modeling the interests and preferences of users toward specific items in e-commerce environments. For example, data analysis techniques are utilized for grouping users that visited, bought or rated similarly the same products (Linden et al., 2003; Su and Khoshgoftaar, 2009). Association rules are used in some cases to relate different products based on their viewing history, e.g., when users view product A and afterwards view product B, then an association rule is created between products A and B indicating a high relationship between the two products (Pierrakos et al., 2003). Accordingly, this information is further utilized by the system to offer recommendations based on the navigation behavior of users.

Apart from modeling the interests of users, adaptive Web 2.0 interactive systems can be developed to accommodate a variety of individual differences, such as cognitive styles. Cognitive styles may correspond to the structure of Web 2.0 environments (Germanakos et al., 2008; Tsianos et al., 2008). Since cognitive styles describe the way users process and organize information, a personalized environment, that is supported by an automated cognitive style-based modeling mechanism, can be adapted at the levels of content selection and structure; the content is essentially either visual or verbal (or auditory), while the manipulation of links can lead to a more analytic and segmented structure, or to a more holistic and cohesive environment. Since the proposed user modeling approach is focused on cognitive style elicitation, we dedicate the next section to existing works on cognitive-based adaptive interactive systems.

2.4. Analysis of existing cognitive-based adaptive interactive systems

Cognitive styles represent the particular set of strengths and preferences that an individual or group of people have in how they take in and process information. By taking into account these preferences and defining specific strategies, empirical research has shown that cognitive styles correlate with performance in a Web-based environment (Wang et al., 2006; Tsianos et al., 2008). Cognitive styles have been defined by Messick (1984) as "consistent individual differences in preferred ways of organizing and processing information and experience, a construct that is different than learning style".

According to Antonietti and Giorgetti (1997) cognitive styles can be measured through the behavioral analysis of users, self-reporting, and through analysis of physiological data. Behavioral data can be obtained by recording the navigation behavior and actions of the users. Self-reports require that people evaluate themselves or answer specially designed questionnaires or psychometric tests. Finally, observations of physiological measures can be taken utilizing bio-metric sensors (e.g., pulse rate, skin conductance).

The literature distinguishes a number of dimensions in which the users' cognitive styles may differ: field-dependent/independent, impulsive/reflective, conceptual/inferential, thematic/relational, analytic/global (Chen and Macredie, 2002; Liu and Ginther, 1999). Regarding the hypermedia information space, among the numerous proposed theories of individual style, a selection of the most appropriate and technologically feasible cognitive styles (those that can be projected on the processes of selection and presentation of Web content and the tailoring of navigational tools) has been studied, such as Riding's Cognitive Style Analysis (CSA) (Verbal/Imagery, and Wholistic/Analytic) (Riding, 1991; Riding and Chema, 1991), Felder/Silverman Index of Learning Styles (ILS) (4 scales: Active vs. Reflective, Sensing vs. Intuitive, Visual vs. Verbal, and Global vs. Sequential) (Felder and Silverman, 1988), Witkin's Field-Dependent and Field-Independent group (Witkin et al., 1977), and Kolb's Learning Styles (Converger, Diverger, Accommodator, and Assimilator) (Kolb and Kolb, 2005), in order to identify how users transform information into knowledge (constructing new cognitive frames).

Since, by its nature, cognitive style influences humans' ability to access and process information, the work on adaptation to user cognitive style was focused more on the presentation of content and the navigational side of adaptive hypermedia systems and adaptive educational systems (Brusilovsky and Millán, 2007). Recently, a considerable amount of research efforts have been focused on modeling and utilizing cognitive factors for personalization in adaptive interactive systems. Several approaches have distinguished users based on their cognitive styles and provided different adaptation effects accordingly (Brusilovsky and Millán, 2007; Botsios and Georgiou, 2009). Table 1 illustrates some noteworthy cognitive-based systems based on the type of individual styles incorporated in the adaptation process and the adaptation effect provided.

In a study, Tsianos et al. (2008) have distinguished Imager and Verbal users, and Wholist and Analyst users based on Riding's Cognitive Style Analysis (Riding, 1991). Each user was provided with adaptive presentation of content and different navigation organization based on their cognitive style. According to Riding's theory (Riding, 1991), users that belong to the Imager class process images more efficiently than text, whereas users that belong to the Verbal class the opposite. Users that belong to the Wholist class have a holistic approach in processing information and have serial navigation behavior whereas users that belong to the Analyst class organize information in pieces of data and have scattered navigation behavior. In a similar approach, Triantafillou et al. (2004) proposed the Adaptive Educational System based on Cognitive Styles system (AES-CS) that distinguished field-dependent and field-independent users based on Witkin et al. (1977) and provided different navigation organization, amount of user control, and navigation support tools for these groups. AES-CS attempted to adapt in terms of both user knowledge and user cognitive style. A field-dependency test was used to classify users in two groups: field-dependent and field-independent users. After that a range of system features were adapted to the identified cognitive style. Field-independent users received an access to the navigation menu to control their navigation. Field-dependent users were only able to proceed through the content sequentially, however they were provided with additional orientation support tools such as a concept map and a path indicator. Depending on their style, the users also received different instructions, feedback, and contextual organizers. Evaluated against a static version of content, the AES-CS (with both kinds of adaptation enabled) demonstrated a significant increase of user performance (Triantafillou et al., 2004).

In their research, Graf and Kinshuk (2009) proposed a framework for adapting course content by distinguishing users to four dimensions based on the Felder/Silverman Index of Learning Styles, i.e., active/reflective, sensing/intuitive, visual/verbal, and sequential/global. The proposed approach estimated the users' learning style through a 44-item questionnaire based on the Felder-Silverman learning style model (Felder and Silverman, 1988). The adaptation effects provided, included the sequence of examples, exercises, and self-assessment tests and determined whether they are presented before the main content, after the main content or at both positions. Another adaptation effect was the number of presented examples and exercises. Moreover, the framework adapted the use of outlines by either presenting them only once before the content or additionally between topics in order to provide students with a better overview. Furthermore, conclusions could be presented either after the content in order to summarize the learned material before applying the knowledge for other tasks (e.g.,

Table 1
Adaptive interactive systems based on individual style theory.

Literature	Individual theory	User modeling technique	Adaptation effects
Triantafillou et al. (2004)	Witkin's Field-Dependent, Field-Independent	Specially designed psychometric test	• Navigation organization • Amount of user control • Navigation support tools
Graf and Kinshuk (2009)	Felder/Silverman Index of Learning Styles	44-item questionnaire	• Sequence of content • Amount of content
Papanikolaou et al. (2003)	Felder/Silverman Index of Learning Styles	Specially designed psychometric test	• Adaptive navigation support (i.e., sequential, global) • Adaptive content presentation (i.e., visual/verbal)
Tsianos et al. (2008)	Riding Cognitive Style Analysis	Specially designed psychometric test	• Adaptive presentation of content • Navigation organization
Carver et al. (1999)	B.S. Solomon's Inventory of Learning Styles	28-item questionnaire	• Adaptive presentation of content
Milosevic et al. (2007)	Kolb learning style model	Navigation behavior	• Different course material sequencing
Karampiperis et al. (2006)	Working Memory and Cognitive Trait Model	Navigation behavior	• Adaptive presentation of content

exercises) or presented at the end of the chapter in order to give students a final summary of the chapter.

In another work, Papanikolaou et al. (2003) investigated users' learning and cognitive style, and preferences during interaction by conducting empirical studies on two educational systems (Flexi-OLM and INSPIRE). The Felder-Silverman Index of Learning Styles questionnaire was used to assess the style of each participant. In Carver et al. (1999) an adaptive hypermedia interface provided adaptive presentation of course material based on the student's learning styles. Students' learning style was extracted by answering a series of 28 questions through an assessment tool developed at North Carolina State University (B.S. Solomon's Inventory of Learning Styles).

In the research of Milosevic et al. (2007), adaptation was provided based on user's preferences, knowledge, goals, navigation history as well as the Kolb's learning style model (Kolb and Kolb, 2005). They suggested that every learning style dimension should get a different course material sequencing.

In Karampiperis et al. (2006), two cognitive factors were utilized, i.e., working memory capacity (Miller, 1956) and inductive reasoning ability (based on the Cognitive Trait Model (Kinshuk and Lin, 2004)), to create adaptivity algorithms. In their work they simulated different learner behavior in navigating a hypermedia space, and measured the selection success of the proposed selection decision model as it is dynamically updated using the simulated learner's navigation steps. The simulation results provided evidence that the proposed selection methodology could dynamically update the internal adaptation logic leading to refine selection decisions (Botsios and Georgiou, 2009).

Taking into consideration the abovementioned works, the next section presents two user studies for eliciting similar groups of users based on their navigation behavior and investigates how these groups may be related to cognitive styles. To the best of the author's knowledge, this is among the first attempts to study the relation between the cognitive style of users and their navigation behavior in an online encyclopedia system, apart from sporadic attempts, like Frias-Martinez et al. (2007) that utilized a number of clustering techniques to understand human behavior and perception in relation with cognitive style, expertise and gender differences of digital library users, Antoniou and Lepouras (2010) that studied the connection between the way people moved in a museum and the way they preferred to approach and process information cognitively, Hsu and Chen (2011) that investigated how learners' cognitive style affect their navigation behavior through data mining techniques as well as analyzed how navigation behavior may influence performance in education environments, and Kinley et al. (2010) that explored the relationships between Web users' searching behavior and their cognitive style. These approaches primarily focus on how individuals use search (e.g., basic or advanced search) and navigation tools (e.g., navigation maps, index of pages, keyword search) and aim to cluster users based on the number of times each feature of the tools are used and further related to the users' cognitive style. Instead, this paper proposes specific user interaction metrics aiming to examine how users navigate (i.e., linearly/non-linearly) based on the sequence of hyperlinks visited in a Web 2.0 environment and further perform clustering techniques on these interaction data to investigate differences in users' navigation behavior and their relation to cognitive styles.

3. Experimental study

The objective of the study is threefold; (i) investigate the relation between cognitive styles and navigation behavior of users, (ii) investigate whether clustering techniques can group users of particular cognitive style using measures obtained from Riding's CSA test, and (iii) evaluate the use of navigation content metrics proposed in this work to find identifiable groups of users that have similar navigation patterns within the group of users that participated in the study. The overarching aim is to gain empirical knowledge for supporting adaptive interactive systems with dynamic user modeling schemes based on users' cognitive style and navigation behavior.

3.1. Cognitive style theory used in the study

Many researchers have developed theories of individual differences in cognitive style, and consequently, argue that individuals have differences in the way they process, organize and remember information (Riding and Cheema, 1991; Felder and Silverman, 1988; Witkin et al., 1977). A defining point in cognitive style research was the work conducted by Riding and Cheema (1991) who surveyed approximately thirty different cognitive styles and concluded that most of them measured two broad cognitive style dimensions: a Wholist/Analyst dimension which refers to how individuals organize information and indicates a preference for information to be structured to get the big picture or the detail, and a Verbal/Imager dimension which refers to how individuals process information and indicates a preference for representing information using pictures or words.

Regarding the Wholist/Analyst dimension, among all the related surveyed cognitive styles, the theory of field-dependency/independency proposed by Witkin (Witkin, 1962; Witkin et al., 1977) is considered one of the most important and highly researched cognitive styles (Rezaei and Katz, 2004; Riding and Cheema, 1991) who distinguished individuals being field-dependent and field-independent. In particular, Witkin describes field-independence as "an analytical, in contrast to global, way

of perceiving which entails a tendency to experience items as discrete from their backgrounds and reflects ability to overcome the influence of an embedding context". For example, when confronted with problems, some individuals are good at extracting things from the context and prefer to handle them in a more analytical way. In contrast, individuals termed as field-dependent cannot abstract an element from its context and are intended to handle problems in a holistic way. In this context, Riding and Cheema proposed the Wholist/Analyst dimension and classify users to the cognitive typologies of Wholist–Intermediate–Analyst. Specifically, users that belong to the Wholist class view a situation and organize information as a whole and are supposed to take a linear approach in hypermedia navigation. Users that belong to the Analyst class view a situation as a collection of parts, stress one or two aspects at a time and are supposed to take a non-linear approach in hypermedia navigation. Users that belong in between the two end points of the Wholist/Analyst scale (i.e., Intermediate) do not differ significantly with regard to information organization.

Regarding the Verbal/Imager dimension, all the related surveyed cognitive styles were developed from the perspective of dual coding theory (Paivio, 1971) which suggests that visual and verbal information is processed and represented differently, and along two distinct cognitive sub-systems in the human mind; the visual and verbal cognitive sub-systems. Each sub-system creates separate representations for information processed which are used to organize incoming information that can be acted upon, stored, and retrieved for subsequent use. From this viewpoint, the Verbal/Imager cognitive style dimension suggests that individuals have differences in the way they process and remember information. In particular, individuals may process verbal information more efficiently than visual information, whilst others the opposite (Peterson et al., 2009; Kozhevnikov, 2007; Riding and Cheema, 1991). Although it is likely that individuals switch strategies depending on the nature of the task, studies have revealed that individuals consistently prefer one or the other strategy (Kozhevnikov, 2007; Riding and Cheema, 1991). Furthermore, ability might affect preference toward a particular strategy in that if a particular mode of processing is more efficient for a person then it is more likely to be preferred (Riding and Cheema, 1991). To this end, Riding and Cheema (1991) have proposed the Verbal/Imager dimension that classifies users to the cognitive typologies of Verbal–Intermediate–Imager. Users that belong to the Verbal class can proportionally process textual and/or auditory content more efficiently than images, whereas users that belong to the Imager class behave in an opposite manner. Users that belong in between the two end points (i.e., Intermediate) do not differ significantly with regard to information processing.

In this context, we have utilized the Wholist/Analyst and Verbal/Imager dimensions of Riding's Cognitive Style Analysis (CSA) (Riding, 1991) and suggest that their implications can be mapped effectively on Web 2.0 environments as described in Table 2, since they consist of distinct scales that respond directly to different aspects of Web 2.0 environments. The CSA implications can provide clear guidelines in the context of Web design (i.e., selecting to present visual/verbal content and structuring information flow in a Wholistic/Analytic manner).

3.1.1. Users' cognitive style elicitation

Users' cognitive styles were elicited by exploiting Riding's CSA test, which measures cognitive style on the two broad Wholist/Analyst and Verbal/Imager dimensions (Riding, 1991), and is considered one of the most credible psychometric tests to elicit cognitive style of users (Riding, 1991; Kinley et al., 2010). In particular, Riding's CSA test comprises of two sub-tests that respectively indicate the position of an individual on each of the Wholist–Analyst and Verbal–Imager dimensions by means of a ratio.

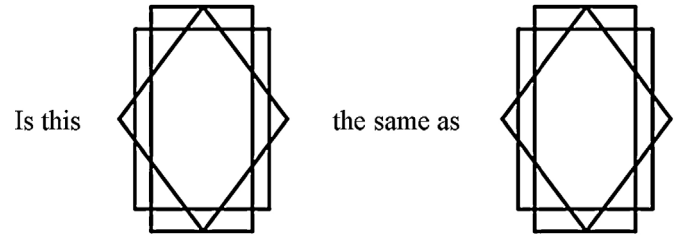


Fig. 1. Example of a Wholist-type stimuli.

The first sub-test assesses the Wholist–Analyst dimension by presenting a series of 40 questions on judging and comparing geometrical figures made up of three basic geometric shapes (i.e., square, rectangle, and triangle). 20 of these questions include wholist-type stimuli that require the participants to compare whether a pair of figures are identical or not (e.g., "Is shape X the same as shape Y?") (Fig. 1), in which half of the items have either the same shape or not. As this task involves judgments about the overall similarity of the two figures, it is assumed that Wholists will respond faster than Analysts. The rest 20 questions include analytic-type stimuli that require the participants to judge whether a single figure is part of another complex figure ("Is shape X contained in shape Y?"), in which half of the items have the shape embedded in the more complex and half of the shapes do not. This task requires the participants to disembed the simple shape from within the complex geometrical figure in order to establish that it is the same as the stimulus shape displayed. It is assumed that Analysts will respond faster at this task. The test records the response time of each given answer to the questions and then uses a three-phase algorithm to determine the participant's cognitive style. The algorithm performs the following steps: (i) calculate the average response time on each of the two sections (20 questions for the wholist-type stimuli, and 20 questions for the analytic-type stimuli) of the CSA test, (ii) calculate the ratio between the average response times on the wholist-type stimuli and analytic-type stimuli, and (iii) associate the value of each subject's Wholist–Analyst ratio with a style category. A low ratio ($\leq$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" (Riding, 1991).

The second sub-test assesses the Verbal–Imager dimension by presenting a series of 48 questions about conceptual category and appearance (i.e., color) to be judged by the participants true or false. 24 statements include verbal-type stimuli that require the participants to compare two objects conceptually (e.g., "Are ski and cricket the same type?"). It is assumed that Verbals respond faster in this type of stimuli since the semantic conceptual category membership is verbally abstract in nature and cannot be represented in visual form. The rest 24 statements include imager-type stimuli that require the participants to compare the color of two objects (e.g., "Are cream and paper the same color?"). In this case, it is assumed that Imagers respond faster to the appearance statements (color) since the objects can be readily represented as mental pictures and the information for the comparison can be obtained directly and rapidly from these images. As in the Wholist–Analyst sub-test, the psychometric test records the response time of each given answer to the questions and then uses a three-phase algorithm to determine the participant's cognitive style. The algorithm performs the following steps: (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

Table 2
Riding's CSA scale mapping to Web 2.0 environments.

CSA scale	Typology	Description	Web implications
Verbal/Imager	Verbal	Process verbal information more efficiently	Web content in text/auditory form
Verbal/Imager	Imager	Process graphical information more efficiently	Web content in graphical, visual representation
Verbal/Imager	Intermediate	In between the Verbal/Imager scale	No significant difference in preference or information processing
Wholist/Analyst	Wholist	Views a situation and organizes information as a whole	Linear approach in Web-site navigation
Wholist/Analyst	Analyst	Views a situation as a collection of parts and only stresses one or two aspects at a time	Non-linear approach in Web-site navigation
Wholist/Analyst	Intermediate	In between the Wholist/Analyst scale	No significant difference in information processing

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" (Riding, 1991).

3.2. Sampling and procedure

A total of 106 individuals participated voluntarily in two user studies carried out within the first week of November 2011 and last week of February 2012. The first experimental study consisted of 22 participants and the second experimental study consisted of 84 participants. All participants were undergraduate students and their age varied from 17 to 25. The students first completed the two sub-tests of Riding's CSA test (Riding, 1991) and then navigated in a reproduced version of a Web 2.0 application; Wikipedia (http://wikipedia.org) which had been enriched to include verbal-based content, i.e., content in textual form without images/visuals in case users belonged to the Verbal cognitive style class (Fig. 2A), and image-based content, i.e., content represented with images/visuals and diagrams in case users belonged to the Imager cognitive style class (Fig. 2B). In this respect, by providing verbal or graphical versions of the same environment, matched to the Verbal/Imager dimension, we aimed to reduce the probability that the content presentation would affect the results of the Wholist/Analyst dimension in which we are primarily interested since it corresponds to the navigation behavior of users.

In the first experimental study, participants were asked to freely navigate through the Wikipedia articles, whereas in the second experimental study participants were assigned 10 problem-based tasks whose answers could be found inside the Wikipedia articles. The first experimental study had the aim to investigate the preference of users between textual and pictorial environments and the results obtained helped in the experimental setup of the second study. The second experimental setup aimed to increase the navigation activity of users and investigate their behavior in solving the problem-based tasks they had been assigned.

3.3. User interactions

Two types of interactions (Fig. 3) were considered for the analysis; (i) *Navigation Menu Interactions*: the interactions of users with a navigation menu of each article in which every hyperlink was connected with a particular section in the article and (ii) *Content Link Interactions*: the interactions of users with hyperlinks within the article that was connected with another article in the system.

The particular interaction types were used in the current study due to the fact that the information behind each hyperlink has close semantic relationship with its previous hyperlink. Accordingly, the usage of the hyperlinks of each interaction type can be used to measure the degree of linear behavior a user has within the Web environment, i.e., whether the user tends to visit successive hyperlinks or not.

3.3.1. Methodology for tracking user interactions

In order to track the users' interactions, all hyperlinks (Navigation Menu and Content Hyperlinks) within each article have been automatically annotated with an attribute, meaningful to the system. In particular, a browser-based logging facility was developed that initially parsed the HTML document, extracting the Navigation Menu hyperlinks and Content hyperlinks, and further annotated each hyperlink in the following manner: Navigation Menu hyperlinks and Content hyperlinks were respectively annotated with "*nav_x_y*" where *x* is the unique identifier of the current article and *y* the unique identifier of the navigation menu hyperlink, and "*content_x_y*" where *x* is the unique identifier of the current arti-

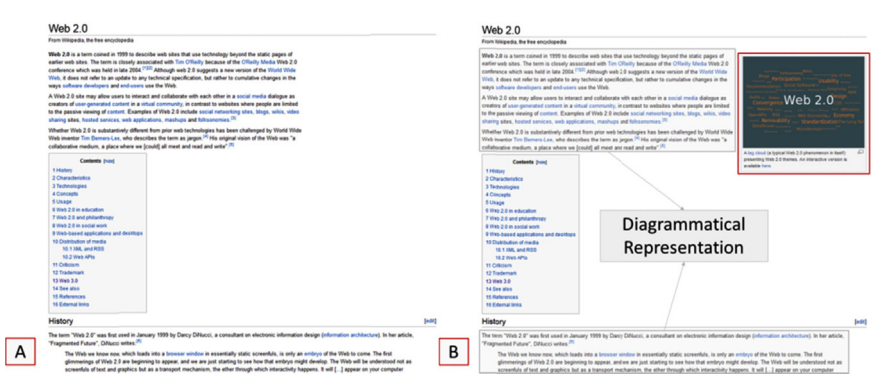


Fig. 2. Verbal- and image-based user interface of the Web-site used in the study.

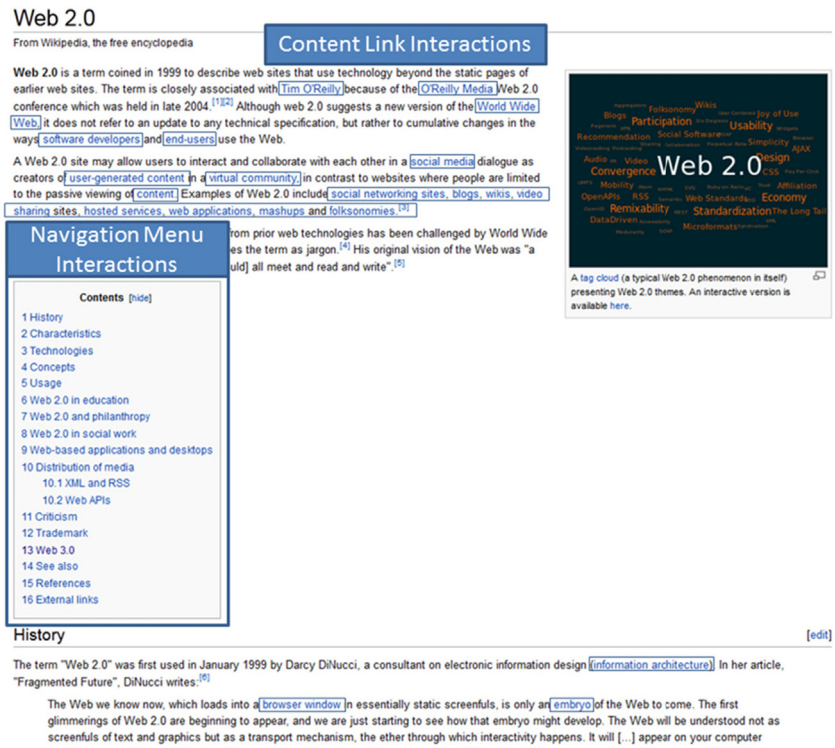


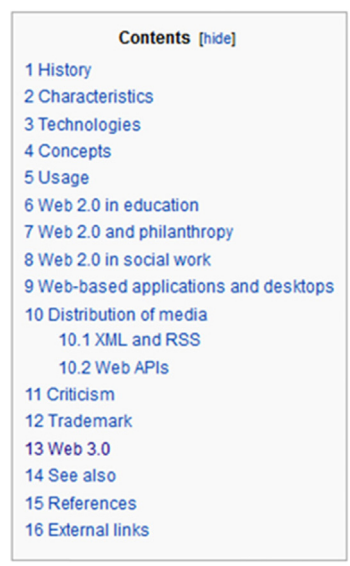
Fig. 3. Types of user interactions.

cle and *y* the unique identifier of the content hyperlink. In both hyperlink types, *y* is used to calculate the distance between the hyperlinks visited by the user. The browser-based logging facility was also responsible to store all users' interactions with each annotated hyperlink in the server's database. Figs. 4 and 5 respectively illustrate the annotations made by the browser-based logging facility.

3.3.2. Definition of user interaction metrics

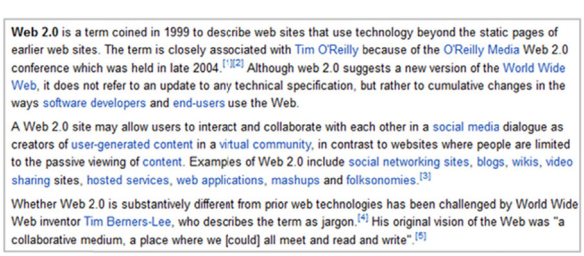
In order to measure the linearity of user interactions with the structured hyperlinks, we created the following metrics that are based on the previously explained hyperlink annotations and types of user interactions:

i. Absolute Distance of Links (ADL), which indicates the similarity between a fully sequential path and the actual path that was traversed by the user, and is calculated by summing the total absolute distance between the links visited by a user.
ii. Average Sequential Links (ASL), which indicates how many sequential links were visited by the user and is calculated by finding the average number of sequential links visited by a user.
iii. Average non-sequential Groups of Links (AGL), which indicates the number of non-sequential links that were visited by the user and is calculated by the average number of non-sequential



Hyperlink	ID
History	nav_1_1
Characteristics	nav_1_2
Technologies	nav_1_3
Concepts	nav_1_4
Usage	nav_1_5
...	nav_x_y

Fig. 4. Navigation menu hyperlink semantic annotations.



Hyperlink	ID
Tim O'Reilly	content_1_1
O'Reilly Media	content_1_2
World Wide Web	content_1_3
software developers	content_1_4
end-users	content_1_5
...	content_x_y

Fig. 5. Content hyperlink semantic annotations.

groups of links visited by a user, if all sequential links are considered to represent one group.

To facilitate our description, assume that a Web-page features a sequence of n consecutive links L_i: $1 \le i \le n$. The goal of the aforementioned metrics is to quantify the user deviations from navigating this Web-page in a sequential manner by visiting each consecutive link L_i (i.e., $L_1, L_2, \ldots$ and finally L_n). Assume that the user has performed the click stream navigation pattern "nav_1_4|nav_1_2|nav_1_3", based on Fig. 4. This navigation pattern translates to "Article 'Web 2.0′, Navigation Menu: 4-2-3", which means that the user visited the "Web 2.0" article and then read the content of the fourth ("Concepts"), second ("Characteristics") and third ("Technologies") hyperlink of the navigation menu. For this particular navigation, as defined above, the metrics are calculated as: $\mathrm{ADL} = (|4-1| + |2-4| + |3-2|)/N = 2$, $\mathrm{ASL} = M/N = 0.333$ and $\mathrm{AGL} = B/N = 0.667$, where N is the number of total links clicked (3 in our case), M is the number of sequential (linear) links used based on the Web-site content (1 in our case) and B is the number of non-sequential groups of links derived from the links used (2 in our case). In our example, B is equal to 2 as for pattern "2–3" the only sequential link clicked was the second and the two non-sequential groups of links were patterns "4" and "2–3". The metrics were also normalized based on the number of user interactions by dividing each variable to the total number of clicks. A main hypothesis for these calculations was that the starting point of all users' interactions in the Web-page is the first link appearing in the navigation menu (e.g., "Web 2.0 – History") since it is the first content the user is interacting with. The abovementioned metrics were used to capture the linearity of the user's navigation on the types of user interactions (Navigation Menu and Content Hyperlinks) with the system, which are described in the next section.

3.4. Analysis of user interactions

The user interaction data analysis performed aims to differentiate users based on their cognitive style dimension (i.e., Wholist–Intermediate–Analyst and Verbal–Intermediate–Imager) and navigation style (i.e., linear and non-linear). The process extracts different groups of users in terms of their responses to the psychometric tests and their Navigation Menu Interactions and Content Link Interactions. Also, an analysis is provided regarding the significance of the interaction data metrics employed which is a complementary target in our analysis.

In particular, the cluster analysis included the following phases: Initially, we defined the optimum number of clusters for each question using two-step clustering (Norusis, 2004) that uses an agglomerative hierarchical clustering method. Single-linkage clustering was used to determine which number of clusters is the optimal in each experiment case and each of the cluster solutions obtained was compared using Schwarz's Bayesian Criterion (BIC). Particularly, in the clustering performed we produced a range of cluster solutions (from 2-cluster solution to 20-cluster solution) 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 represent the most well-separated clusters. The ideal solution in this case was when the density inside clusters is high and between clusters is low. 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. In the final step of the process, we used the results of two-step clustering to analyze the distribution and significance of the metrics. Details on the k-means and two-step clustering processes can be respectively found in Hartigan (1975) and Norusis (2004).

In the following sections we present in detail the two experimental studies performed and for each we present the results of the k-means and two-step clustering process.

3.5. Experimental Study I – results and discussion

This section presents and analyses the results obtained from the first experimental study. The following analyses were performed: (i) k-means clustering on the answers of users to the psychometric test, (ii) k-means clustering on the navigation pattern of users (interaction metrics) in the online encyclopedia system, and (iii) two-step clustering in order to analyze the metric's significance and distribution.

Table 3 summarizes the cognitive-based groups of users based on the measures obtained from the psychometric test. As already explained, the ratios between the response times of the users to specific stimuli were computed and then a simple rule-based mechanism was applied on the ratio to assign the user to a cognitive typology. An example of a rule is "if a user had a ratio of response times less or equal than 1.02 then the user is assigned to the Wholist group" (Riding, 1991). Accordingly, based on the ratios of all users and the rule-based mechanism, cognitive-based groups of users were created as summarized in Table 3. The first three columns of Table 3 and the last three columns refer to the Wholist/Analyst and Verbal/Imager dimensions respectively. Regarding the Wholist/Analyst dimension, Intermediates outnumbered the Wholists (7) and Analysts (3) with a total of 12 users, whereas the Verbal/Imager dimension, Imagers outnumbered the Intermediates (5) and Imagers (6) with a total of 11 users.

Furthermore, we wanted to investigate whether specific clustering mechanisms could create cognitive-based groups of users based on the measures obtained from the psychometric test. Based on the clustering performed using the results of the psychometric test we observe that the users are clustered in three groups whose range of cognitive style is reported in Table 4. The table also presents the total number of users included in each cluster. The figures show that the users are clearly distinguished based on their cognitive profile and that they cover the whole range of the scale suggested by Riding (1991). For example, in the Wholist/Analyst dimension one of the clusters contains users with cognitive style ratio in the [1.776, 1.853] interval which is in line to Riding's Analyst scale (i.e., >1.35) and in the Verbal/Imager dimension the clustered users' cognitive style ratio in the [0.832, 0.941] interval is again in line to Riding's Verbal scale (i.e., $\le$0.98). Overall, the number of users in each cluster is very close to Riding's grouping, i.e., in the Wholist/Analyst range only $\pm$1 users, and in the Verbal/Imager range case, only $\pm$2 users were differently classified. Another observation is that the range of users' Wholist/Analyst and Verbal/Imager dimensions vary slightly (shown in higher figures) compared to Riding's scale. These findings indicate that the k-means clustering technique can group users of particular cognitive styles utilizing measures obtained from psychometric tests. The technique has provided encouraging results and justifies further utilization.

The next cluster analysis involves visit path analysis of the users' interactions with the system (i.e., Navigation Menu Interactions and Content Link Interactions) by using the three proposed clustering metrics which measure the linearity of the users' navigation behavior (i.e., how linear the navigation behavior is). The statistics of the actual values for each metric (i.e., ADL, ASL and AGL) which measure the linearity of the users' navigation menu visit path in a different manner and as formed in clusters are reported in Table 5.

The figures indicate that the users of the same cluster had similar navigation behavior (i.e., were navigating in a linear/non-linear

Table 3
Cognitive style ratio based on Riding's psychometric test.

Group	Users	Wholist/Analyst range	Group	Users	Verbal/Imager range
Wholist	7	[0.786, 1.000]	Verbal	11	[1.121, 1.248]
Intermediate	12	[1.030, 1.337]	Intermediate	5	[0.998, 1.040]
Analyst	3	[1.424, 1.853]	Imager	6	[0.832, 0.962]

Table 4
Cognitive style ratio of clustered users based on Riding's psychometric test.

Cluster	Users	Wholist/Analyst range	Users	Verbal/Imager range
1	8	[0.786, 1.030]	6	[1.121, 1.248]
2	12	[1.099, 1.424]	7	[0.958, 1.040]
3	2	[1.776, 1.853]	9	[0.832, 0.941]

manner). Specifically, all three metrics grouped consistently the users in clusters characterized by linear/non-linear users based on their navigation behavior; Cluster 1 contains users with linear navigation behavior, e.g., based on ADL, 10 users had a range between 0 and 1.667 that indicates a linear approach of navigation, whereas in Cluster 2, based on ADL, 2 users had a range between 3 and 4 that indicates a non-linear approach. Remember that the ADL metric measures the distance between each link visited by the users. This finding encourages the use of the above metrics for further investigation. The distinct groups of users formed can be used to adapt the design of Web applications to accommodate the needs of each group of users navigated in a Web environment.

Fig. 6(a)–(c) shows the variance of the metrics ADL, ASL and AGL using box plots in relation to the clusters formed in the first three cases of Table 5. The box plots show that each metric can distinguish effectively the users in two distinct clusters (significantly different values and mean values are observed in each cluster case).

Table 6 presents the obtained ranges of the users' cognitive-based profiles in each cluster based on the users' interactions with the navigation menu and Fig. 7 the membership degree of each user in the clusters formed based on the metrics. For example, the cognitive style of the users grouped based on their interactions with the navigation menu in the first cluster is between the values 0.827 and 1.853 regarding the Wholist/Analyst dimension and between the values 0.856 and 1.248 regarding the Verbal/Imager dimension. Accordingly, the respective cognitive typologies of users of each cluster are variant (range 0.827–1.853 regarding the Wholist/Analyst dimension) indicating that currently, no safe conclusions can be drawn whether cognitive styles of users correlate with the navigation behavior of users (i.e., linear or non-linear).

Fig. 7 shows for each cluster formed by using the metrics ADL, ASL and AGL what type of cognitive typologies appear on the left based on Riding's CSA and on the right based on a modified CSA. The modified CSA was used to further distinguish some Intermediate users to Light Analysts or Light Wholists. In particular, in the case of Wholists the rule "if a user has a ratio of response times higher than 1.02 but lower than 1.15 then the user is assigned to the Light Wholist group" was applied, whereas in the case of Analysts the rule "if a user has a ratio of response times higher than 1.15 but lower or equal as 1.35 then the user is assigned to the Light Analyst group". The main observation is that in each cluster mixed types of cognitive style users exist based on Riding's CSA. However, using the modified CSA the majority of users in the first cluster that had linear navigation behavior are Wholists and Light Wholists. In this respect, such a finding suggests that applying a fuzzy rule in the cognitive style classification process of users reveals more clear relation between users' navigation behavior and cognitive style.

Furthermore, the same cluster analysis was performed on the interactions of the users with content links (i.e., hyperlinks of the article that were connected with other articles of the Web environment). Table 7 suggests that the proposed metrics group users in clusters based on their linear/non-linear navigation behavior. In particular, all three metrics have grouped users that had linear navigation behavior in Cluster 1 and users with non-linear navigation behavior in Cluster 2 based on the Content Link Interactions of users.

The box plots illustrated in Fig. 8 show that the two clusters formed in using the Content Link Interactions taking into consideration the three proposed metrics are significantly different in terms of values and mean values.

Furthermore, Table 8 presents the obtained ranges of the users' cognitive profiles in each cluster based on their interactions with the hyperlinks inside the content and Fig. 9 the membership degree of each user in the cluster based on the metrics. The results summarized in Table 8 indicate that the clusters created based on the sequence of interactions with the content links had users with variant cognitive typologies suggesting that no strong relation seems to exist between the cognitive styles of users and their navigation behavior. A possible interpretation of these results can be based on the fact that some successive content links may not have a close semantic relationship and thus not clearly reflected on the navigation behavior of users (linear or non-linear).

For both navigation menu and Content Link Interactions, two-step clustering was also performed using the combination of all three metrics, primarily to obtain each metrics' significance, visualize the distribution of each proposed metric used for clustering and confirm their ability to cluster users in significantly different clusters. The distance was measured using the log-likelihood distance measure between each case and its classification center. The clustering results of two-step clustering for the Navigation Link and Content Link Interactions are illustrated in Fig. 10. The dark shaded features in the third row indicate which metrics are the most significant in obtaining the clusters. The distribution of each metric in each cluster is also illustrated in the same row. The darker and the

Table 5
Descriptive statistics of clusters ($k = 2$) per metric based on Navigation Menu Interactions.

Metric	Cluster 1				Cluster 2			
	Users	Range	Mean	Stdev	Users	Range	Mean	Stdev
ADL	10	[0.000, 1.667]	1.283	0.489	2	[3.000, 4.000]	3.500	0.707
ASL	2	[0.583, 1.000]	0.792	0.295	10	[0.000, 0.286]	0.084	0.107
AGL	4	[0.500, 0.750]	0.601	0.123	8	[0.800, 1.000]	0.903	0.083

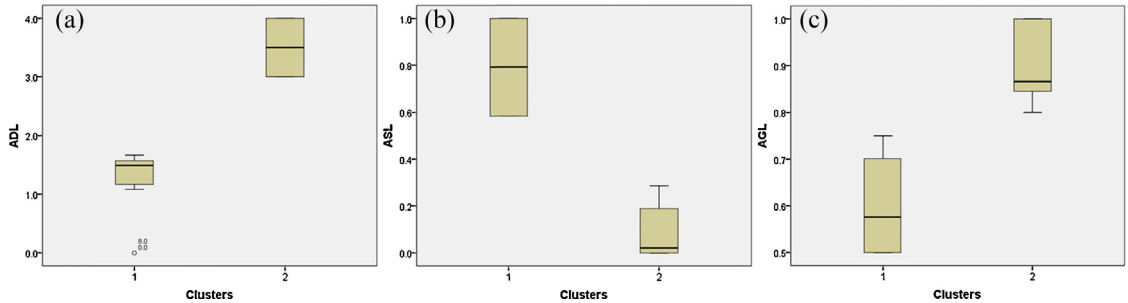


Fig. 6. Box plots of clusters formed for Navigation Menu Interactions using (a) ADL, (b) ASL and (c) AGL metrics.

Table 6
Cognitive style ratio of clustered users ($k = 2$) based on Navigation Menu Interactions.

Metric	Range of Cluster 1		Range of Cluster 2	
	Wholist/Analyst	Verbal/Imager	Wholist/Analyst	Verbal/Imager
ADL	[0.827, 1.278]	[0.856, 1.248]	[0.849, 1.853]	[0.998, 1.126]
ASL	[1.221, 1.278]	[0.856, 0.962]	[0.827, 1.853]	[0.874, 1.248]
AGL	[0.827, 1.221]	[0.856, 1.124]	[0.849, 1.853]	[0.874, 1.248]

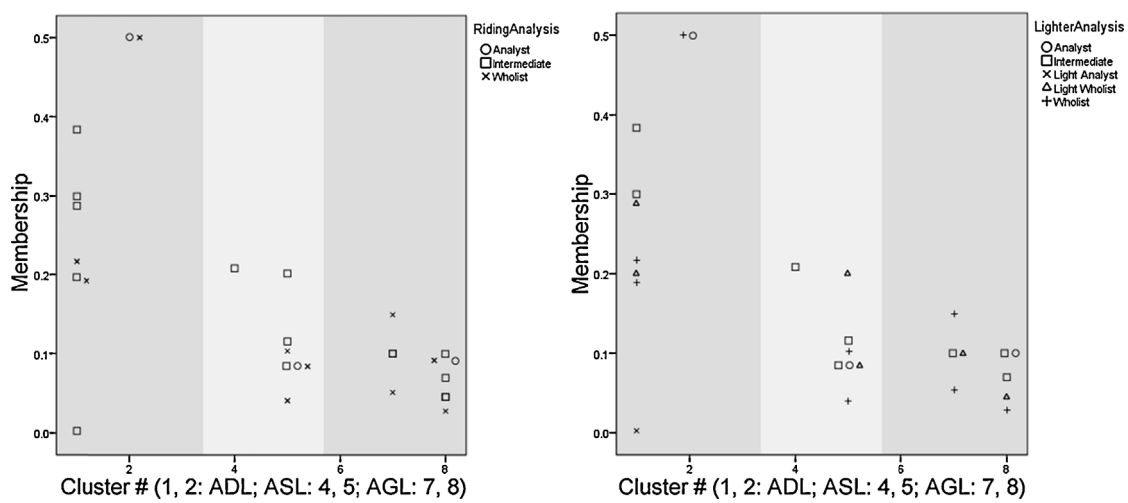


Fig. 7. Membership in each cluster for metrics ADL, ASL and AGL for the Navigation Menu Interactions using (a) Riding's CSA and (b) modified CSA.

Table 7
Descriptive statistics of clusters ($k = 2$) per Metric based on Content Link Interactions.

Metric	Cluster 1				Cluster 2			
	Users	Range	Mean	Stdev	Users	Range	Mean	Stdev
ADL	9	[0.000, 1.625]	1.171	0.502	12	[1.778, 3.333]	2.258	0.467
ASL	6	[0.500, 1.000]	0.616	0.199	15	[0.000, 0.412]	0.247	0.133
AGL	8	[0.500, 0.750]	0.639	0.103	13	[0.786, 1.000]	0.925	0.088

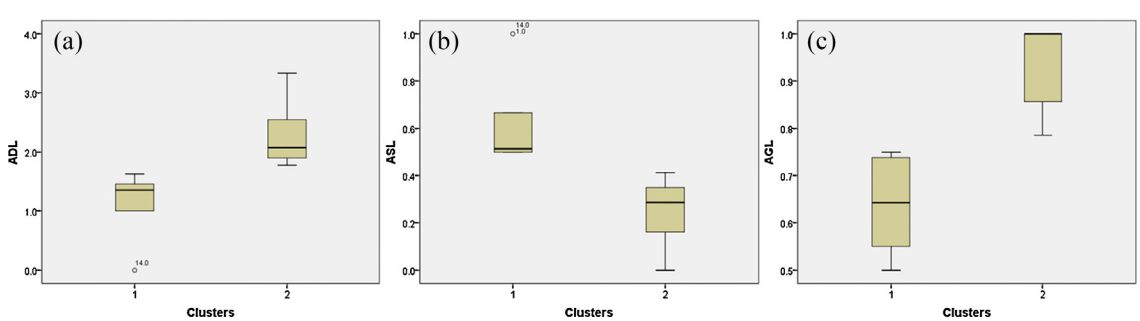


Fig. 8. Box plots of clusters formed for Content Link Interactions using (a) ADL, (b) ASL and (c) AGL metrics.

Table 8
Cognitive style ratio of clustered users ($k = 2$) based on Content Link Interactions.

Metric	Range of Cluster 1		Range of Cluster 2	
	Wholist/Analyst	Verbal/Imager	Wholist/Analyst	Verbal/Imager
ADL	[0.819, 1.776]	[0.861, 1.248]	[0.786, 1.853]	[0.832, 1.154]
ASL	[0.819, 1.248]	[0.861, 1.121]	[0.786, 1.853]	[0.832, 1.248]
AGL	[0.819, 1.776]	[0.885, 1.248]	[0.786, 1.853]	[0.832, 1.128]

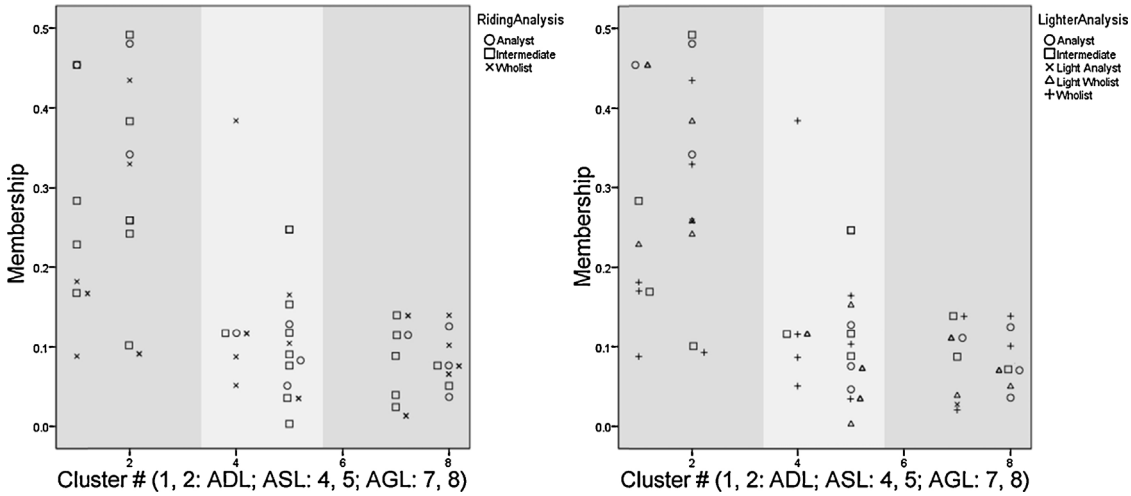


Fig. 9. Membership in each cluster for metrics ADL, ASL and AGL for the Content Link Interactions using (a) Riding's CSA and (b) modified CSA.

lighter distribution indicate the cluster and the overall data distribution respectively. The last row shows the distribution of cognitive style typologies of the users in each cluster based on Riding's CSA. In the case of Navigation Menu Interactions, the ADL metric is the most significant variable in obtaining two clusters ($k = 2$). Furthermore, the majority of users were grouped in Cluster 1 given that most of them were Wholists and Light Wholists and had linear navigation behavior. In addition, some users were not included in any of the two clusters since no sufficient data existed about their interactions with the Navigation Menu. In the Content Link Interactions case, the ADL metric, as well as the rest metrics are important. The distribution allows us to observe for the ADL metric in the Content Link Interactions for instance, a lower proportion of values for the users in Cluster 1 and a higher proportion for the users in Cluster 2 exist. In this respect, the ADL metric could be used in a personalization engine that would implicitly identify the navigation behavior of the Web-site visitors and further correlated with their cognitive style to decide which adaptation effects to communicate to the user interfaces.

In sum, no safe conclusion can be drawn whether cognitive styles can relate to the navigation behavior of users (in regard with linear/non-linear approach). Nevertheless, our addition of two new

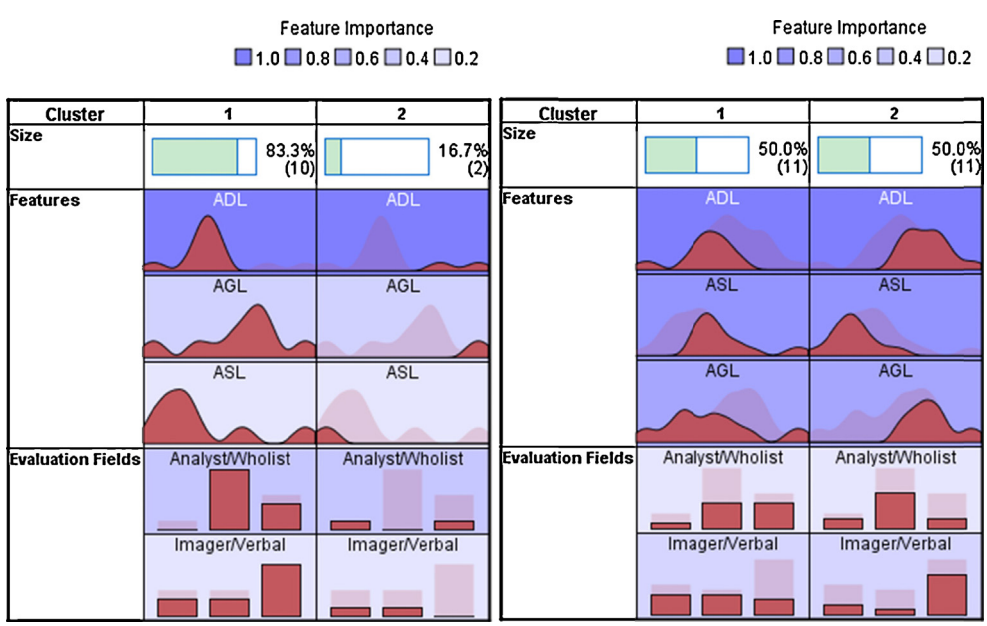


Fig. 10. Clustering results of the (a) Navigation Menu Interactions and (b) Content Link Interactions of Experimental Study I.

Table 9
Cognitive style ratio of users based on Riding's psychometric test.

Group	Users	Wholist/Analyst range	Group	Users	Verbal/Imager range
Wholist	30	[0.543, 1.002]	Verbal	41	[0.599, 0.980]
Intermediate	29	[1.046, 1.347]	Intermediate	22	[0.980, 1.084]
Analyst	25	[1.377, 3.170]	Imager	21	[1.093, 1.489]

classes to the Riding's CSA (i.e., users that were classified as Intermediates and their ratio of answers was closer to the threshold of the Wholist or Analyst class, were classified as Light Wholists or Light Analysts respectively), has shown that many Wholists and many former Intermediates (and now Light Wholists) had linear navigation according to the metrics proposed. Specifically, consistently the majority of users in Cluster 1 formed using all metrics, which had linear navigation behavior, were classified as Wholists and Light Wholists. Such finding is in line with theory (Riding and Cheema, 1991), that users belonging to the Wholist class tend to have a more controlled and serial approach during navigation. The results have also shown that *k*-means clustering technique can group users of particular cognitive styles using measures obtained from psychometric tests and that the proposed metrics can be used to classify users in distinctive groups based on their navigation behavior (i.e., linear/non-linear).

3.6. Experimental Study II – results and discussion

The second study aimed to examine a sample of users with increased interaction in the Web 2.0 environment. The navigation behavior and the cognitive style typologies were investigated in a similar context but when dealing with problem-based tasks. As previously explained, the users were asked a set of questions for which the answers could be found in the articles and by navigating through the Web content. Similarly with Experimental Study I, the following analyses were performed: (i) *k*-means clustering on the answers of users to the psychometric test, (ii) *k*-means clustering on the navigation pattern of users (interaction metrics) in the online encyclopedia system, and (iii) two-step clustering in order to analyze the metric's significance and distribution.

Table 9 summarizes the cognitive-based groups of users involved in the experiment based on the measures obtained from the psychometric test. Regarding the Wholist/Analyst dimension, users were distributed along the cognitive style range with 30 Wholists, 29 Intermediates and 25 Analysts. On the other hand, in the Verbal/Imager dimension, Verbals outnumbered the Intermediates (22) and Imagers (21) with a total of 41 users.

Table 10 summarizes the cognitive-based groups of users based on the same clustering mechanisms applied in Experimental Study I. The figures report whether these mechanisms could create cognitive-based groups of users based on the measures obtained from the psychometric test. Regarding the Verbal/Imager dimension, the results indicate that the users are similarly distinguished based on their cognitive profile as in Table 9. However, a difference between the groupings of Verbals and Intermediates of the two methods can be observed. In particular, the rule-based mechanisms applied, based on Riding's suggested thresholds (Riding, 1991), groups 41 Verbals and 22 Intermediates whereas the clustering mechanism groups 30 Verbals and 35 Intermediates. This may be due to the fact that the clustering performed does not strictly apply Riding's thresholds to decide on the typologies but decides to form the clusters based on the metrics of the psychometric test. Thus, the clustering mechanism used in this study optimally recalibrates the thresholds to form clusters based on the intrinsic characteristics of the sample under investigation. Therefore, for example in the Verbal/Imager dimension the clustering performed has grouped 13 more users as Intermediates which when further analyzed (based on their cognitive ratio) were close to the Verbals' and Imagers' threshold (i.e., may be considered as Light Verbals and Imagers). Hence, clustering mechanisms could be used as an alternative method for highlighting differences in cognitive style of users and capturing the fuzziness of the information available (i.e., cognitive style ratios) to extract optimum groups of users.

Table 10
Cognitive style ratio of clustered users based on Riding's psychometric test.

Cluster	Users	Wholist/Analyst range	Users	Verbal/Imager range
1	73	[0.543, 1.660]	30	[0.599, 0.898]
2	7	[1.534, 3.111]	35	[0.921, 1.102]
3	4	[1.619, 3.170]	19	[1.121, 1.489]

Tables 11–14 summarize the results based on the cluster analysis of the users' interactions with the system (i.e., Navigation Menu Interactions and Content Link Interactions). Table 11 summarizes the clusters created based on the proposed metrics using the interactions of users with the Navigation Menu. Similarly to Experimental Study I, the results indicate that the users of the same cluster group had similar navigation behavior (i.e., linear/non-linear navigation). In particular, Cluster 1 contains users with linear navigation behavior, e.g., based on ADL 44 users had a range between 0 and

Table 11
Descriptive statistics of clusters ($k = 2$) per metric based on Navigation Menu Interactions.

Metric	Cluster 1				Cluster 2			
	Users	Range	Mean	Stdev	Users	Range	Mean	Stdev
ADL	44	[0.000, 1.348]	1.109	0.280	34	[1.375, 3.000]	1.587	0.290
ASL	2	[1.000, 1.000]	1.000	0.000	76	[0.000, 0.250]	0.077	0.068
AGL	59	[0.200, 0.619]	0.443	0.107	19	[0.636, 1.000]	0.810	0.144

Table 12
Cognitive style ratio of clustered users ($k = 2$) based on Navigation Menu Interactions.

Metric	Range of Cluster 1		Range of Cluster 2	
	Wholist/Analyst	Verbal/Imager	Wholist/Analyst	Verbal/Imager
ADL	[0.545, 3.111]	[0.707, 1.489]	[0.543, 3.170]	[0.599, 1.442]
ASL	[1.061, 1.119]	[0.932, 0.985]	[0.543, 3.170]	[0.599, 1.489]
AGL	[0.543, 1.534]	[0.734, 1.442]	[0.545, 3.170]	[0.599, 1.489]

Table 13
Descriptive statistics of clusters ($k = 2$) per metric based on Content Link Interactions.

Metric	Cluster 1				Cluster 2			
	Users	Range	Mean	Stdev	Users	Range	Mean	Stdev
ADL	40	[0.000, 1.833]	1.078	0.509	18	[2.000, 4.000]	2.738	0.642
ASL	15	[0.571, 1.000]	0.811	0.190	43	[0.000, 0.500]	0.214	0.210
AGL	25	[0.286, 0.714]	0.546	0.134	33	[0.750, 1.000]	0.982	0.062

Table 14
Cognitive style ratio of clustered users ($k = 2$) based on Content Link Interactions.

Metric	Range of Cluster 1		Range of Cluster 2	
	Wholist/Analyst	Verbal/Imager	Wholist/Analyst	Verbal/Imager
ADL	[0.731, 3.170]	[0.599, 1.242]	[0.543, 3.111]	[0.702, 1.442]
ASL	[0.543, 3.170]	[0.599, 1.242]	[0.684, 3.111]	[0.702, 1.442]
AGL	[0.684, 3.111]	[0.702, 1.442]	[0.543, 3.170]	[0.599, 1.242]

1.348 that indicates a linear approach of navigation, whereas in Cluster 2, based on ADL 34 users had a range between 1.375 and 3 that indicates a non-linear approach. Based on ASL, the resulting clustering produced Cluster 1 with 2 users and Cluster 2 with 76 users. This indicates that at least for the current sample the average number of sequential (linear) links visited by the users were not different in order to discriminate the users in two clusters. In fact, most users (76) followed a non-linear navigation behavior according to the sequence of links used.

Fig. 11 also illustrates the clusters formed using the three metrics for each clustering case of Table 11 through box plots. It is evident that only the clustering using the AGL metric distinguishes clearly two clusters (i.e., Clusters 1 and 2) since they present high variances in terms of values and mean values.

Similarly to the results obtained from the Experimental Setup I, Tables 12 and 14, suggest that no absolute correlation exists between cognitive styles of users and their navigation behavior (i.e., linear vs. non-linear). However, the box plots in Fig. 13 show that the ADL and the ASL metrics can distinguish users of particular navigation behavior based on their Content Link Interactions.

Fig. 12 presents the cognitive styles of users in clusters formed using the Navigation Menu Interactions taking into consideration Riding's CSA and the modified CSA as explained previously. The modified CSA typically assigned Intermediate users in the Light Wholist or Light Analyst classes in an average rate of 49% and 20% respectively across the clustering performed for all the metrics, with the majority having linear navigation behavior. In this respect, similarly to Experimental Study I, applying a fuzzy rule in the cognitive style classification process of users revealed more clear relation between users' navigation behavior and cognitive style since users tending to be Wholists had linear navigation behavior.

Finally, Table 13 indicates that the proposed metrics group users that had linear navigation behavior in Cluster 1 and users with non-linear navigation behavior in Cluster 2 based on the Content Link Interactions of users.

Based on the box plots of Fig. 13 we reach to the same conclusion as in the previous experimental study, that the two clusters formed using the Content Link Interactions and by taking into consideration the three metrics are significantly different in terms of values and mean value. The box plots clearly show that Cluster 1 and Cluster 2 in all metrics contain users following linear and non-linear navigation behavior respectively using Content Link Interactions.

Fig. 14 illustrates the cognitive styles of the users in each cluster formed using the Content Link Interactions based on Riding's CSA and on the modified CSA. The main observation from this analysis is that the modified CSA now assigns Intermediate users in the Light Wholist or Light Analyst classes in an average rate of 47% and 26% respectively across the clustering performed for all the metrics. In this case, no clear relationship was observed between cognitive styles and navigation behavior of users since cognitive styles varied across all clusters, with some cases however indicating that based on the modified CSA analysis, the number of Wholist users was slightly bigger in the cluster with users having followed a linear navigation.

In addition, the results using two-step clustering for the Navigation Menu and Content Link Interactions are presented in Fig. 15 using the combination of all three metrics, with the aim to obtain each metrics' significance, visualize the distribution of each proposed metric used for clustering and confirm their ability to cluster users in significantly different clusters. As in Experimental Study I, the distance was measured using the log-likelihood distance measure between each case and its classification center. The most dark shaded feature in the third row indicates the most significant variable in obtaining the two clusters ($k = 2$) formed. The distribution of each metric in each cluster is also illustrated in this row. The darker and the lighter distribution indicate the cluster and the

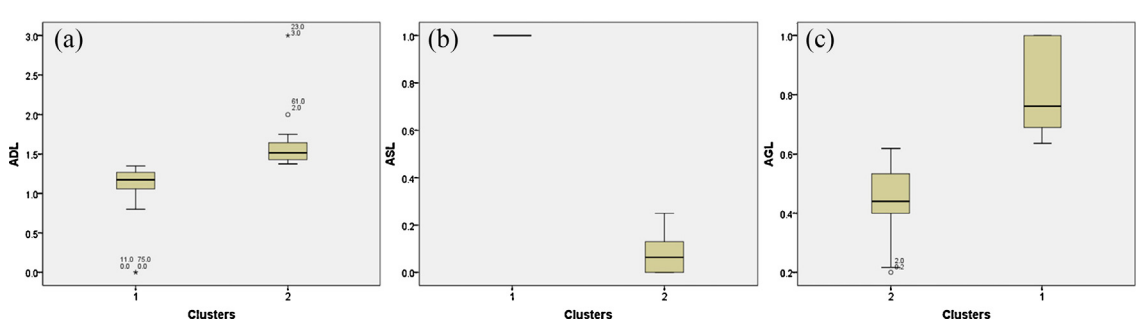


Fig. 11. Box plots of clusters formed for Navigation Menu Interactions using (a) ADL, (b) ASL and (c) AGL metrics.

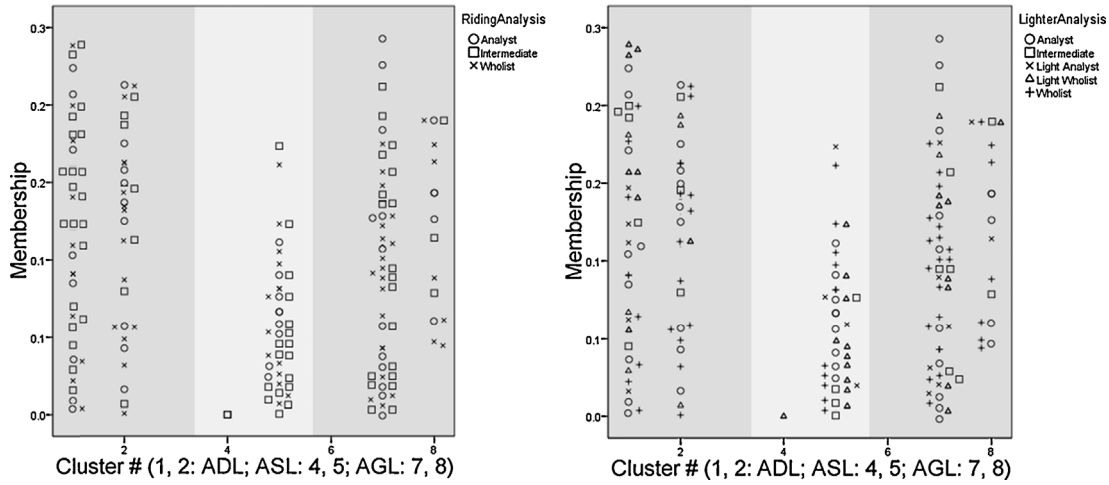


Fig. 12. Membership in each cluster for the metrics ADL, ASL and AGL for the Navigation Menu Interactions of Experimental Study II using (a) Riding's CSA and (b) modified CSA.

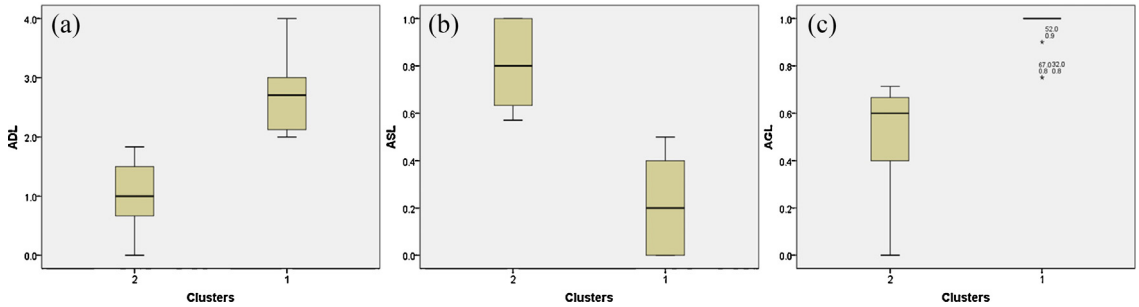


Fig. 13. Box plots of clusters formed for Content Link Interactions using (a) ADL, (b) ASL and (c) AGL metrics.

overall data distribution respectively while the last row shows the distribution of cognitive style typologies of the users in each cluster. In the Navigation Menu Interactions case the dark shaded feature in the third row indicates that the ASL metric is the most significant variable in obtaining these two clusters ($k = 2$). The distribution of each metric in each cluster (illustrated in the same row) indicates that the majority of users had a sequential approach on the navigation menu, and were therefore grouped in Cluster 1. In addition, the majority of these users tended to be Wholists (after applying the fuzzy rule on the cognitive style ratios) revealing a relationship between cognitive style and navigation behavior of users. Regarding the Content Link Interactions, the AGL and ASL metrics are the most significant variables whose distribution is left skewed in Cluster 1 and right skewed in Cluster 2.

In sum, results revealed that *k*-means clustering can group users of particular cognitive styles using measures obtained from psychometric tests and that the user interaction metrics proposed could be used to classify users in distinctive groups based on their navigation behavior (i.e., linear/non-linear). Also, results indicate that differences exist in navigation style between users (linear/non-linear),

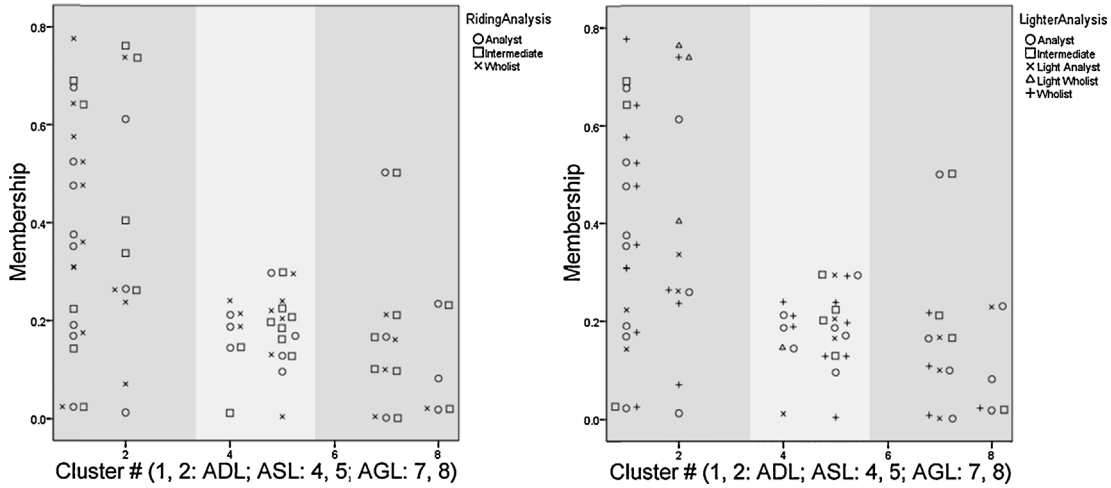


Fig. 14. Membership in each cluster for the metrics ADL, ASL and AGL for the Content Link Interactions of Experimental Study II.

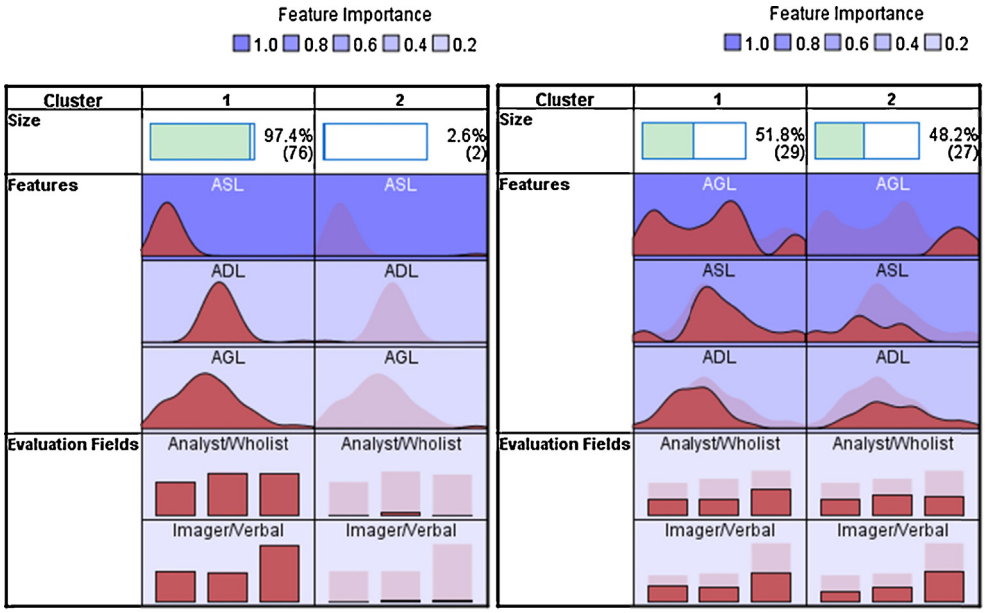


Fig. 15. Clustering results of the (a) Navigation Menu Interactions and (b) Content Link Interactions of Experimental Study II.

however, no safe conclusion can be drawn whether the navigation style can be strongly related to cognitive styles of users. Nevertheless, by applying the fuzzy rule to the cognitive style ratio (i.e., classifying Intermediate users to Light Wholists or Light Analysts), has shown that users being Wholists, or that tended being Wholists (Light Wholists) had linear navigation according to the metrics proposed. Such an observation suggests that fuzzy rules could be used as an alternative to the existing strict rule-based method for classifying users in a particular cognitive style group, and thus handling the classification uncertainty of cognitive style which is driven by the dynamicity of human nature.

4. Conclusions

The overarching aim of this work was to increase our understanding on navigation behavior of users in relation to their cognitive style by studying their interaction within Web 2.0 environments utilizing specific clustering techniques. Accordingly, specific navigation metrics have been proposed and utilized by a clustering technique, with the aim to identify groups of users that have similar navigation behavior and investigate the relation to their cognitive style. Such a finding could provide a promising direction toward the identification of adaptation rules, based on the abovementioned relationship, for a more user-centric interface design. A practical implication of this work could be based on a personalization engine that would implicitly identify the cognitive style of Web-site visitors based on their navigation behavior and further feed an adaptation engine with the users' models providing different adaptation effects. Adaptation effects based on different cognitive style of users could for example group all the links of an article in a floating menu in case of a Wholist user, thus, supporting linear navigation behavior, or create a tabbed navigation menu for enabling/disabling the corresponding section of an article in case of an Analyst user. Furthermore, the paper also investigated whether *k*-means clustering can group users of particular cognitive style using measures obtained from psychometric tests.

The clustering of users based on the proposed navigation metrics have shown promising results since the clustering technique grouped consistently the users based on their navigation behavior (i.e., linear/non-linear). Such results indicate that individuals navigate differently in terms of linearity among hyperlinks that have close semantic relationship. Furthermore, investigating the relationship between cognitive styles (based on Riding's CSA thresholds) and the navigation pattern followed by each user revealed that users of the same cluster group, although had similar navigation behavior (i.e., linear/non-linear), their respective cognitive style was variant. This is maybe due to potential weaknesses of the reproduced version of Wikipedia at a conceptual level which needs to be further studied or due to the strict rule-based algorithm applied on the users' cognitive style ratios which might not be able to handle the uncertainty of cognitive styles driven by the dynamicity and complexity of human nature. On the other hand, a modified cognitive style analysis applied on the clustered users revealed that users with linear navigation behavior (belonging to the same cluster) were commonly classified as Wholists or Light Wholists (previously classified as Intermediates according to Riding's CSA), which is in line with theory that users of the Wholist class tend to have a controlled and serial approach to navigation (Riding and Cheema, 1991). Accordingly, the proposed metrics could be used by Web 2.0 providers for tracking the navigation behavior of users and implicitly classifying users to a particular cognitive style group for further utilization by decision making and adaptation mechanisms. For example, an adaptive Web 2.0 environment could decide which type of navigation support could be provided to the users based on the degree of linearity of their navigation and cognitive style, e.g., users with linear navigation behavior and holistic cognitive style could be provided with supportive navigation support tools stressing linear navigation through the hyperlinks with the aim to increase the efficiency and effectiveness of task completion and overall improve the users' experiences.

Finally, results have also shown that *k*-means clustering could be applied effectively on the data extracted from cognitive style elicitation tests by assigning the users into a cognitive-based cluster. Results revealed that clustering mechanisms can be used as an alternative method for forming groups of users with particular cognitive style, handling the uncertainty and fuzziness of the information available (e.g., ratio) to extract optimum groups of users, rather than using a strict rule-based mechanism that highlights differences in cognitive style based on specific thresholds.

The limitations of the current work are related to the difficulty to obtain full control in such complex Web environments, the small sample of users participating in the study, the number

of clusters used in the analysis (for example in larger samples perhaps a higher number of clusters could be found), the selection of clustering method, the effect of outliers and the order of cases analyzed. We have addressed some of these threats by examining various cluster sizes and clustering mechanisms and reaching to a conclusion which ones to use. On the optimum number of clusters we decided to extract number of clusters based on the cognitive profiles of the users. In addition, since we also had a moderately sized set of samples this meant that the selection of *k*-means clustering was a reasonable choice. The results of *k*-means clustering may depend on the order of the users using the online environment and thus we have had arranged the samples in a random order to address this threat. In addition, the fact that the *k*-means clustering algorithm is sensitive to the initial randomly selected cluster centers we have eliminated this threat by repeating the algorithm execution several times.

A future research prospect is to conduct further studies investigating the relation of cognitive styles of users with other types of navigation behavior than the ones investigated in this work (i.e., linear/non-linear) as well as capture and study the users' interactions with other Web objects (e.g., drop down lists, drag and drop tools, Web forms). In addition, carrying out a single assessment of users' cognitive style might not fully justify the users' classification into specific cognitive style groups since individuals might be influenced by other circumstances over time such as emotions, urgency, etc. In this respect, further studies need to be conducted in order to reach to more concrete conclusions about the effect of cognitive style of users on their navigation behavior and the adaptation of Web 2.0 environments. Furthermore, we plan to investigate the effect of users' navigation behavior during collaborative and sharing activities as well as evaluate the proposed approach and metrics in other Web environments (e.g., social networks, collaboration platforms). Another future activity might include automatically predicting the cognitive style of users utilizing their navigation behavior using artificial intelligence techniques. Since content and functionality of Web 2.0 environments is either presented visually, verbally or auditory, and users may have specific navigation behavior, e.g. linear or non-linear navigation behavior, this work can be useful for assisting adaptive Web 2.0 environments to organize and present information and functionalities in an adaptive format to diverse user groups.

Acknowledgements

The work is co-funded by the PersonaWeb project under the Cyprus Research Promotion Foundation (ΤΠΕ/ΠΛΗΡΟ/0311(BIE)/10), and the EU projects Co-LIVING (60-61700-98-009) and SocialRobot (285870).

References

Antonietti, A., Giorgetti, M., 1997. The Verbalizer–Visualizer Questionnaire: a review. Perceptual and Motor Skills 86 (1), 227–239.

Antoniou, A., Lepouras, G., 2010. Modeling visitors' profiles: a study to investigate adaptation aspects for museum learning technologies. Journal on Computing and Cultural Heritage 3 (2), 1–19.

Botsios, S., Georgiou, D., 2009. Recent contributions to a generic architecture design that supports learning objects interoperability. Distance Educational Technologies 7 (4), 64–85.

Bra, P.D., Aerts, A., Houben, G.J., Wu, H., 2000. Making general-purpose adaptive hypermedia work. In: Proceedings of WebNet World Conference on the WWW and Internet, pp. 117–123.

Brusilovsky, P., 2001. Adaptive hypermedia. User Modeling and User-Adapted Interaction 11 (1–2), 87–110.

Brusilovsky, P., Millán, E., 2007. User models for adaptive hypermedia and adaptive educational systems. In: Kobsa, A., Nejdl, W. (Eds.), The Adaptive Web. LNCS, vol. 4321. Springer, Heidelberg, pp. 3–53.

Brusilovsky, P., Kobsa, A., Nejdl, W., 2007. The Adaptive Web: Methods and Strategies of Web Personalization. Springer, Heidelberg.

Carver Jr., C.A., Howard, R.A., Lane, W.D., 1999. Enhancing student learning through hypermedia courseware and incorporation of student learning styles. IEEE Transactions on Education 42 (1), 33–38.

Castellano, G., Torsello, M.A., 2008. Categorization of web users by fuzzy clustering. In: Proceedings of International Conference on Knowledge-Based Intelligent Information and Engineering Systems, Springer, pp. 222–229.

Castellano, G., Fanelli, A.M., Mencar, C., Torsello, M.A., 2007. Similarity-based fuzzy clustering for user profiling. In: Proceedings of International Conference on Web Intelligence and Intelligent Agent Technology Workshop. IEEE/WIC/ACM, pp. 75–78.

Chen, S.Y., Macredie, R.D., 2002. Cognitive styles and hypermedia navigation: development of a learning model. Journal of the American Society for Information Science and Technology 53 (1), 3–15.

Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R., Robbins, D., 2003. Stuff I've seen: a system for personal information retrieval and re-use. In: Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval. ACM Press, pp. 72–79.

Eirinaki, M., Vazirgiannis, M., 2003. Web mining for web personalization. ACM Transactions on Internet Technology 3 (1), 1–27.

Felder, R., Silverman, L., 1988. Learning and teaching styles in engineering education. Engineering Education 78, 674–681.

Frasincar, F., Barna, P., Houben, G., Fiala, Z., 2004. Adaptation and reuse in designing web information systems. In: Proceedings of IEEE Conference on Information Technology: Coding and Computing. IEEE Computer Society, pp. 291–387.

Frias-Martinez, E., Magoulas, G.D., Chen, S.Y., Macredie, R.D., 2005. Modeling human behavior in user-adaptive systems: recent advances using soft computing technique. Expert Systems with Applications 29 (2), 320–329.

Frias-Martinez, E., Chen, S.Y., Macredie, R.D., Liu, X., 2007. The role of human factors in stereotyping behavior and perception of digital library users: a robust clustering approach. User Modeling and User-Adapted Interaction 17 (3), 305–337.

Fu, Y., Sandhu, K., Shih, M.Y.,1999. Clustering of web users based on access patterns. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Springer.

Gauch, S., Speretta, M., Chandramouli, A., Micarelli, A., 2007. User profiles for personalized information access. In: Kobsa, A., Nejdl, W. (Eds.), The Adaptive Web. LNCS, vol. 4321. Springer, Heidelberg, pp. 54–89.

Germanakos, P., Tsianos, N., Lekkas, Z., Mourlas, C., Samaras, G., 2008. Realizing comprehensive user profile as the core element of adaptive and personalized communication environments and systems. The Computer Journal 52 (7), 749–770.

Google Desktop, 2012. http://desktop.google.com (accessed 21.02.12).

Graf, S., Kinshuk, 2009. Advanced adaptivity in learning management systems by considering learning styles. In: Proceedings of International Workshop on Social and Personal Computing for Web-Supported Learning Communities, pp. 235–238.

Hartigan, J.A., 1975. Clustering Algorithms. John Wiley and Sons, New York.

Houben, G., Frasincar, F., Barna, P., Vdovjak, R., 2004. Modeling user input and hypermedia dynamics in hera. In: Proceedings of International Conference on Web Engineering, pp. 60–73.

Hsu, Y., Chen, S., 2011. Associating learners' cognitive style with their navigation behaviors: a data-mining approach. In: Proceedings of International Conference on Human–Computer Interaction: Users and Applications. Springer-Verlag, pp. 27–34.

iGoogle, 2012. http://www.google.com/ig (accessed 03.03.12).

Jawaheer, G., Szomszor, M., Kostkova, P., 2010. Comparison of implicit and explicit feedback from an online music recommendation service. In: Proceedings of International Workshop on Information Heterogeneity and Fusion in Recommender Systems. ACM Press, pp. 47–51.

Joshi, A., Joshi, K., 2000. On mining web access logs. In: Proceedings of ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery. ACM Press, pp. 63–69.

Karampiperis, Lin, P., Sampson, T., Kinshuk, D.G., 2006. Adaptive cognitive-based selection of learning objects. Innovations in Education and Teaching International 43 (2), 121–135.

Kelly, D., Teevan, J., 2003. Implicit feedback for inferring user preference: a bibliography. ACM SIGIR Forum 37 (2), 18–28.

Kinley, K., Tjondronegoro, D., Partridge, H., 2010. Web searching interaction model based on user cognitive styles. In: Proceedings of the Conference of the Computer–Human Interaction Special Interest Group of Australia on Computer–Human Interaction. ACM Press, pp. 340–343.

Kinshuk, Lin, T., 2004. Cognitive profiling towards formal adaptive technologies in web-based learning communities. International Journal of Web Based Communities 1 (1), 103–108.

Kolb, A., Kolb, D., 2005. The Kolb Learning Style Inventory – Version 3.1 Experience Based Learning Systems. Technical Specifications. Experience Based Learning Systems.

Kozhevnikov, M., 2007. Cognitive styles in the context of modern psychology: toward an integrated framework of cognitive style. Psychological Bulletin 133 (3), 464–481.

Linden, G., Smith, B., York, J., 2003. Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Computing 7 (1), 76–80.

Liu, Y., Ginther, D., 1999. Cognitive styles and distance education. Journal of Distance Learning Administration 2 (3).

Messick, S., 1984. The nature of cognitive styles: problems and promises in educational research. Educational Psychologist 1, 59–74.

Miller, G.A., 1956. The magical number seven plus or minus two: some limits on our capacity for processing information. Psychological Review 63 (2), 81–97.
Milosevic, D., Brkovic, M., Debevc, M., Krneta, R., 2007. Adaptive learning by using SCOs metadata. Interdisciplinary Journal of Knowledge and Learning Objects 3, 163–174.
Mobasher, B., 2007. Data mining for web personalization. In: Kobsa, A., Nejdl, W. (Eds.), The Adaptive Web. LNCS, vol. 4321. Springer, Heidelberg, pp. 90–135.
Nasraoui, O., Soliman, M., Saka, E., Badia, A., Germain, R., 2008. A web usage mining framework for mining evolving user profiles in dynamic web sites. IEEE Transactions on Knowledge and Data Engineering 20 (2), 202–215.
Norusis, M., 2004. SPSS 13.0 Statistical Procedures Companion. Prentice Hall Inc., Upper Saddle-River, NJ.
Paivio, A., 1971. Imagery and Verbal Processes. Holt, Rinehart & Winston, New York.
Papanikolaou, K.A., Grigoriadou, M., Kornilakis, H., Magoulas, G.D., 2003. Personalising the interaction in a web-based educational hypermedia system: the case of INSPIRE. User Modeling and User-Adapted Interaction 13 (3), 213–267.
Perkowitz, M., Etzioni, O., 2000. Adaptive web sites. Communications of the ACM 43 (8), 152–158.
Peterson, E., Rayner, S., Armstrong, S., 2009. Researching the psychology of cognitive style and learning style: is there really a future? Learning and Individual Differences 19 (4), 518–523.
Pierrakos, D., Paliouras, G., Papatheodorou, C., Spyropoulos, C.D., 2003. Web usage mining as a tool for personalization: a survey. User Modeling and User-Adapted Interaction 13 (4), 311–372.
Rezaei, A.R., Katz, L., 2004. Evaluation of the reliability and validity of the cognitive styles analysis. Personality and Individual Differences 26, 1317–1327.
Riding, R., 1991. Cognitive Style Analysis – Research Administration. Learning and Training Technology.
Riding, R., Cheema, I., 1991. Cognitive styles – an overview and integration. Educational Psychology 11 (3–4), 193–215.
Schneider-Hufschmidt, M., Kühme, T., Malinowski, U., 1993. Adaptive user interfaces: principles and practice. In: Human Factors in Information Technology. Amsterdam, North-Holland.
Su, X., Khoshgoftaar, T., 2009. A survey of collaborative filtering techniques. Advances in Artificial Intelligence 4, 19 p.
Trajkova, J., Gauch, S., 2004. Improving ontology-based user profiles. In: Proceedings of RIAO 2004, pp. 380–389.
Triantafillou, E., Pomportis, A., Demetriadis, S., Georgiadou, E., 2004. The value of adaptivity based on cognitive style: an empirical study. Journal of Educational Technology 35 (1), 95–106.
Tsianos, N., Lekkas, Z., Germanakos, P., Mourlas, C., Samaras, G.,2008. User-centered profiling on the basis of cognitive and emotional characteristics: an empirical study. In: Proceedings of International Conference on Adaptive Hypermedia and Adaptive Web-based Systems, LNCS, vol. 5149. Springer, Heidelberg, pp. 214–223.
Wærn, A., 2004. User involvement in automatic filtering: an experimental study. User Modeling and User-Adaptive Interaction 14 (2–3), 201–237.
Wang, K.H., Wang, T.H., Wang, W.L., Huang, S.C., 2006. Learning styles and formative assessment strategy: enhancing student achievement in web-based learning. Journal of Computer Assisted Learning 22 (3), 207–217.
Witkin, H.A., 1962. Psychological Differentiation. Studies of Development. New York, Wiley.
Witkin, H.A., Moore, C.A., Goodenough, D.R., Cox, P.W., 1977. Field-dependent and field-independent cognitive styles and their educational implications. Review of Educational Research 47 (1), 1–64.

Marios Belk is currently a Ph.D. Candidate at the Department of Computer Science, University of Cyprus. He obtained his B.Sc. and M.Sc. in Computer Science from the same department. His research interests lie in the area of Adaptive Interactive Systems, User Modeling, Web Technologies and the Semantic Web. His M.Sc. thesis contributed for the publication of several research articles in journals, book chapters and major International conferences in which he was co-author in twenty of them. He also actively participates in numerous National and EU funded research projects.

Dr. Efi Papatheocharous is currently a visiting Lecturer at the Department of Computer Science, University of Cyprus. Her Ph.D. studies were focused primarily in Software Engineering with specific research interest in Computational Intelligence and Software Cost Estimation. She obtained her M.Sc. from the University of Manchester in Advanced Computer Science with ICT Management, in 2005, her B.Sc. in Computer Science from the Computer Science Department, University of Cyprus, in 2004. She has been collaborating in a number of research projects at a National and European level and co-authored more than 20 papers published in books, journals and presented in international conferences.

Dr. Panagiotis Germanakos is currently an Assistant Professor in the Department of Management and MIS of the School of Business, University of Nicosia. He obtained his Ph.D. from the University of Athens in 2008 and his M.Sc. in International Marketing Management from the Leeds University Business School in 1999. He has over 65 publications including book, journal, and conference contributions and is co-editor of three books. Furthermore, he actively participates in numerous national and EU funded projects that mainly focus on the analysis, design and development of open interoperable integrated wireless/mobile and personalized technological infrastructures and systems.

Dr. George Samaras is a Professor at the Department of Computer Science, University of Cyprus. He received a Ph.D. and M.Sc. in Computer Science from Rensselaer Polytechnic Institute, USA, in 1989 and 1985 respectively. He obtained his B.Sc. degree in Mathematics at the University of Athens, Greece (1982). His broad research interests span the areas of Mobile and Wireless Computing, Sensor Networks, Database Systems, Transaction Processing, Commit Protocols and Resource Recovery, Personalization and Context-based Systems, WWW Information Retrieval. He has published more than 150 articles in conferences and journals and co-authored two books and a number of book chapters.