Personalization Systems and Processes Review based on a Predetermined User Interface Categorization

Panagiotis GERMANAKOS[1], Constantinos MOURLAS[1], Christoforos PANAYIOTOU[2], George SAMARAS[2]
[1]*Faculty of Communication and Media Studies, National & Kapodistrian University of Athens, 5 Stadiou Str., PC 105-62, Athens, Hellas*
{pgerman, mourlas}@media.uoa.gr
[2]*Computer Science Department, University of Cyprus, CY-1678 Nicosia, Cyprus*
{cs95gp1, cssamara}@cs.ucy.ac.cy

Abstract. The plethora of information and services as well as the complicated nature of most Web structures intensify the navigational difficulties that arise when users navigate their way through this large information space. Personalised services that are highly sensitive to the immediate environment and the goals of the user can alleviate the orientation difficulties experienced by the relatively diverse user population. Personalization is the process of customizing a Web site to users' specific requirements using the knowledge acquired from the analysis of users' navigational behaviour in correlation with other information collected in the Web context as well as other related individual intellectual, mental, emotional and social-context elements. With the emergence of the wireless and mobile technologies, these needs have been extended to requirements for continuous interaction in the "*anytime, anywhere and anyhow*" context. Since the personalization dimensions for both desktop and mobile users are relatively diverse and since a standard definition has not yet been developed, this paper will make an extensive review of the personalization considerations, categories and paradigms, comparing and contrasting processes and techniques in an attempt to establish a framework that will act as the base for the formulation of a more concrete personalization definition, platform and device independent.

1. Introduction

Since 1994, the Internet has emerged as a fundamental information and communication medium that has generated extensive enthusiasm. The Internet has been adopted by the mass market more quickly than any other technology over the past century and is currently providing an electronic connection between progressive businesses and millions of customers and potential customers whose age, education, occupation, interest, and income demographics are excellent for sales.

The explosive growth in the size and use of the World Wide Web as well as the complicated nature of most Web structures result in orientation difficulties, as users often lose sight of the goal of their inquiry, look for stimulating rather than informative material, or even use the navigational features unwisely. As the e-business sector is rapidly evolving, the need for such Web structures that satisfy the heterogeneous needs of its users is becoming more and more evident.

To alleviate such navigational difficulties, researchers have put huge amounts of effort to identify the peculiarities of each user group and analyze and design methodologies and systems that could deliver up-to-date personalized information, with regards to products or services. Since to date there has not been a concrete definition of personalization, the many solutions offering personalisation features meet an abstract common goal: to provide users with what they want or need without expecting them to ask for it explicitly [1]. Further consideration and analysis of parameters and contexts such as users intellectuality, mental capabilities, socio-psychological factors, emotional states and attention grapping strategies, that could affect the apt collection of users' customization requirements offering in return the best adaptive environments to their preferences and demands should be extensively investigated.

With the emergence of wireless and mobile technologies, new communication platforms and devices, apart from PC-based Internet access, are now emerging making the delivery of information available through a variety of media. Inevitably, this increases user requirements which are now focused upon an "*anytime, anywhere and anyhow*" basis. Nowadays, researchers and practitioners not only have to deal with the challenges of adapting to the heterogeneous user needs and with user environment issues such as current location and time [2], but also have to face numerous considerations with respect to multi-channel delivery of the applications. To this end, personalization techniques exploit Artificial Intelligence, agent-based, and real-time paradigms to give presentation and navigation solutions to the growing user demands.

The aim of this paper is to identify a general personalization framework that would give a concrete definition of personalisation and that could be used as a base for the employment of other more theoretical dimensions. In particular, this paper will report on the personalization considerations related to the desktop and mobile user, presenting the technological challenges and constraints that personalisation methodologies have to incorporate. An extensive investigation on personalization will be presented, paying particular attention to link personalization, content personalization, context personalization, authorized personalization, and humanized personalization, the five user interface categories.

The paper is structured in 6 sections. Section 2 gives an overview of the personalization considerations with regards to the desktop and mobile user. Section 3 describes user profiling for personalization. In section 4 the five personalization categories are presented and section 5 emphasizes on the personalization technologies.

2. Personalization Considerations in the Context of Desktop and Mobile Users

Internet users usually look for information either by Web browsing, that is by navigating from page to page along Web links, or by searching with a search engine [3]. The impersonal organization of information on the Web has some negative consequences for users. In recent survey, Pitkow and Kehoe found that the main problems Internet users encounter when using the Web include slow network or connection speeds; not being able to find specific pages, even after they have been found before; not being able to manage or organize retrieved information; and not being able to visualize where they have been [4].

The science behind personalization has undergone tremendous changes in recent years with the basic goal of personalization systems to remain the same, to provide users with what they want or need without requiring them to ask for it explicitly. Personalization is the provision to the individual of tailored products, services, information or information relating to products or services. Since it is a multi-dimensional and complicated area (covering also, recommendation systems, customization, adaptive Web sites, Artificial

Intelligence) there has not so far been written a universal definition that would cover all its theoretical and technological areas. Nevertheless, most of the definitions have been given to personalization [4, 5, 6, 7, 8, 9, 10] are converging to the objective that is expressed on the basis of delivering to a group of individuals relevant information that is retrieved, transformed, and / or deduced from information sources in the format and layout as well as specified time intervals. More technically, it includes the modeling of Web objects (products, and pages) and subjects (users), their categorization, locating possible similarities between them and determining the required set of actions for personalization. On the other hand, many argue that for the actual meaning of personalization, not only personalised information needs but also emotional or mental needs, caused by external influences, should be taken into account.

Personalization could be realized in one of two ways: (a) Web sites that require users to register and provide information about their interests, and (b) Web sites that only require the registration of users so that they can be identified [11]. The main motivation points for personalization can be divided into those that are primarily to facilitate the *work* and those that are primarily to accommodate *social* requirements. The former motivational subcategory contains the categories of enabling access to information content, accommodating work goals, and accommodating individual differences, while the latter eliciting an emotional response and expressing identity [10].

2.1 The Mobility Emergence – Personalization Challenges and Constraints

The needs of mobile users differ significantly from those of desktop users. Getting personalized information "*anytime, anywhere and anyhow*" is not an easy task. Researchers and practitioners have to take into account new adaptivity axes along which the personalized design of mobile e-services would be built. Such applications should be characterized by flexibility, accessibility, quality and security in a ubiquitous interoperable manner.

As the number and variety of the new emerging channels (networks and devices) is quite large, with differing capabilities and limitations, the design and implementation complexity is rising significantly. With regards to networks, some of the newly addressed concerns are low bandwidth and unreliable connectivity. As for the mobile devices, some of the newly addressed concerns are small size, lack of processing power, limited interface and data entry, small memory and storage space, and high latency. User interfaces must be friendlier enabling active involvement (information acquisition), giving the control to the user (system controllability), providing easy means of navigation and orientation (navigation), tolerating users' errors and support error system-based and context-oriented correction of users' errors, and finally enabling customization of multi-media and multi-modal user interfaces to particular users' needs [11, 12, 13, 14, 15, 16, 15, 2, 18]. To overcome those problems intelligent techniques have to be implemented that will enable the development of an open Adaptive Mobile Web [12]. The fundamental characteristics of these techniques should include directness, high connectivity speed, reliability, availability, context-awareness, broadband connection, interoperability, transparency and scalability, expandability, effectiveness, efficiency, personalization, security and privacy [19, 20, 21, 22].

2.2 The Personalization Problem

Indisputably, the user population is not homogeneous. To be able to deliver quality knowledge, systems should be tailored to the needs of individual users providing them personalized and adapted information. Although one-to-one service provision may be a

functionality of the distant future, user segmentation is a very valuable step towards that direction. User segmentation means that the user population is subdivided, into more or less homogeneous, mutually exclusive subsets of users who share common user profile characteristics. The subdivisions could be based on:

- *Demographic characteristics* (i.e. age, gender, urban or rural based, region)
- *Socio-economic characteristics* (i.e. income, class, sector, number of employees, volume of business, channel access)
- *Psychographic characteristics* (i.e. life style, values, sensitivity to new trends)
- *Individual physical and psychological characteristics* (i.e. disabilities, attitude, loyalty).

User characteristics and needs, determining user segmentation and thus provision of the adjustable information delivery, differ according to the circumstances and they change over time [2, 18, 19, 23].

The issue of personalization is a complex one with many aspects that need to be analyzed. Some of these issues become even more complicated once viewed from a mobile user's perspective, when wireless communication media and mobile device constraints are involved. Such issues include, but are not limited to:

- *What content to present to the user*. How to decide what to show, using user profiles, using the user history to predict future needs etc. When using user profiles the need for (1) storing the interests of the user in a format that is easy to be used, be updated or *moved*, and (2) relating interests and items based on a semantics level (e.g., the theme interest of "flowers" is related to "florists" or even fertile producers) must be addressed.
- *How to show the content to the user*. Many users want to see the same things presented in a different format. In the wireless environment this also relates to the specific characteristics of the mobile device.
- *How to ensure the user's privacy*. Every personalizing system acquires information about the habits of each user. This leads to privacy concerns as well as legal issues [21]. It could also leads to lack of user trust and could result in the failure of the system due to avoidance of its use.
- *How to create a global personalization scheme*. The user does not mind if a set of sites can be personalized but could very well be annoyed when at each one of them they have to repeat the personalization process. This is especially annoying and cumbersome for the user on the move carrying a resource poor mobile device.

These major issues of personalization could be summarized in the following phrase: "What, how and for everything" [2]. There are many approaches to address these issues of personalization but usually, each one is focused upon a specific area, i.e. whether this is profile creation, machine learning and pattern matching, data and Web mining or personalized navigation.

3. User Profiling for Web Personalization

One of the key technical issues in developing personalization applications is the problem of how to construct accurate and comprehensive profiles of individual users and how these can be used to identify a user and describe the user behaviour, especially if they are moving [24]. According to Merriam- Webster dictionary the term profile means "a representation of something in outline". User profile can be thought of as being "a set of data representing the significant features of the user" [25]. Its objective is the creation of an information base that contains the preferences, characteristics, and activities of the user. A user profile can be

built from a set of keywords that describe the user preferred interest areas compared against information items.

User profiling is becoming more and more important with the introduction of the heterogeneous devices used, especially when published services provide customized views on information. When considering different users there must be a distinction of their needs. This means that they may probably want better personalized services to save them time and trouble, or that they may require different level of entrance according to different levels of group awareness, or that they may probably need multiple profiles according to their status. Therefore, capabilities like bandwidth, displays, and text-writing must be taken into account when developing such services.

User profiling can either be *static*, when it contains information that rarely or never changes (e.g. demographic information), or *dynamic*, when the data change frequently. Such information is obtained either *explicitly*, using online registration forms and questionnaires resulting in static user profiles, or *implicitly*, by recording the navigational behaviour and / or the preferences of each user. In the case of implicit acquisition of user data, each user can either be regarded as a member of group and take up an aggregate user profile or be addressed individually and take up an individual user profile. The data used for constructing a user profile could be distinguished into the following:

(a) The *Data Model* which could be classified into:
   a. *demographic* which describes who the user is and,
   b. *transactional* which describes what the user does.

(b) The *Profile Model* which could be further classified into:
   a. the *factual profile,* containing specific facts about the user derived from transactional data, including the demographic data, such as "the favourite beer of customer X is BeerA" and,
   b. the *behavioural profile,* modelling the behaviour of the user using conjunctive rules, such as association or classification rules. The use of rules in profiles provides an intuitive, declarative and modular way to describe user behaviour [24].

Additionally, in the case of a mobile user, by user needs it is implied both, the *thematic preferences* (i.e., the traditional notion of profile) as well as the characteristics of their personal device called *"device profile"*. Therefore, here, adaptive personalization is concerned with the negotiation of user requirements and device abilities.

As Web developers regard personalization as the best way to filter out unnecessary or irrelevant information for their users, some argue on issues like personalization may restrict the extent and the variety of information users receive, that people often do not have well-defined preferences, the need to answer detailed questions to personalize their Web pages, that the recommendation process is a black box for end users and so on [10].

4. Web Personalization Categories

The previous sections have been described the main high level considerations of Web personalization. However, in order to have a more comprehensive insight for its context, it is necessary to classify personalization in categories. These include:

Link Personalization. This strategy involves selecting the links that are more relevant to the user, changing the original navigation space by reducing or improving the relationships between nodes. E-commerce applications use link personalization to recommend items based on the clients' buying history or some categorization of clients based on ratings and opinions. Link personalization is widely used in Amazon.com to link the home page with recommendations, new releases, shopping groups, etc. [29, 30]

Content Personalization. When content becomes personalized, user interface can present different information for different users providing substantive information in a node, other than link anchors. Most of the content personalization research is relative to text and hypertext personalization and can be further classified into two types:

(a) *Node structure customization (personalization)*, usually appears in those sites that filter the information that is relevant for the user, showing only sections and details in which the user may be interested. The user may explicitly indicate their preferences, or these may be inferred (semi-) automatically either from the user profile or navigation activity. For example, in my.yahoo.com or in www.mycnn.com users choose a set of "modules" and further personalize those modules by choosing a set of attributes of the module to be perceived. Some "automatic" customization may occur based on location information (e.g by using the zip code of the user to select local to the user sport events). The outcome of these applications is that the user should be able to "build" their own page.

(b) *Node content customization (personalization)*, occurs when different users perceive different values for the same node attribute; this kind of content personalization is finer grained than structure personalization. A good example can be found in online stores that give customers special discounts according to their buying history (in this case the attribute price of item is personalized) [29, 30].

Context Personalization. Personalizing navigational contexts is critical when the same information (node) can be reached in different situations [29]. A navigational context is a set of nodes that usually share some property. For example in a Conference Paper Review Application, it is possible to access papers etc. Notice that one paper may appear in different sets and that different users may have different access restrictions according to their role in the Review application. Context personalization can also be adapted to the preferences of the learner and semantics of the learner's current environment. One sub-category of context personalization is *terminal adaptivity*. That is adapting information to the characteristics of a device. It is applied on the mobile devices to satisfy learner's demand for "learning as you go". **Terminal Personalization** occurs on a per session basis. Personalization can be achieved by applying many axes of adaptation effecting both the navigational structure and appearance of the learning experience. It involves the tailoring of a resource to the current environment of the learner [31, 19].

Authorized Personalization. In the personalized user interface, different users have different roles and therefore they might have different access authorizations. For example, in an academic application, instructors and students have different tasks to perform. Instructors want to access their class materials, such as upload, edit their class syllabus and give students' grades etc. On the other hand, students want to access the interface to find out their current GPA, their enrollment status, and their course work status etc.

Humanized Personalization. Bonnie Kaplan and Ramesh Farzanfar presented and studied an intelligent interactive telephone system (Telephone-Linked Care (TLC)) that provided information whether they were talking to a machine or to a person during TKC relationships with the TLC system [32]. If the dimension of the "emotional user interface" could be involved, it will be a huge step towards a concrete and universal definition of Web personalization. Unquestionably, this category of personalization still needs to be explored, with an extensive use of Artificial Intelligence technologies, since there are a lot of ambiguity and technical obstacles at present [33].

5. Web Personalization Paradigms Comparison

Web personalization can be defined as the process of customizing the content and structure of a Web site to the specific individual needs of each user taking advantage of the user's navigational behavior. The steps of the Web personalization process include: (1) The collection of Web data, (2) the modelling and categorization of these data (pre-processing phase), (3) the analysis of the collected data, and (4) the determination of the actions that should be performed. The technologies that are employed in order to implement these processing phases are distinguished to:

Content-based filtering. Systems that are implementing these kinds of techniques are solely based on individual users' preferences. The system tracks each user's behavior and recommends items that are similar to items the user liked in the past. It is based on description analysis of the items rated by the user and correlations between the content of these items and user's preferences. It is an alternative paradigm that has been used mainly in the context of recommending items such as books, Web pages, news, etc. for which informative content descriptors exist [34, 35, 36]. This technique is primarily characterized by two weaknesses, *content Limitations* and *over-Specialization*. There are content limitations like IR methods that can only be applied to a few kinds of content, such as text and image, and the extent aspects can only capture certain aspects of the content. On the other hand content-based recommendation systems provide recommendations merely based on user profiles, therefore, users have no chance of exploring new items that are not similar to those items included in their profiles and thus leading to over-specialization. Consequently, some more drawbacks that have been identified in time are [37, 36, 38]:

(a) Search-based models build keyword, category, and author indexes offline, but fail to provide recommendations with interesting, targeted titles. They also scale poorly for customers with numerous purchases and ratings.
(b) User input may be subjective and prone to bias.
(c) Explicit (and non-binary) user ratings may not be available.
(d) Profiles may be static and can become outdated quickly.
(e) May miss other semantic relationships among objects.

At this point it would be noteworthy to mention a complementary technique of Content-based filtering, namely **Social Information filtering**. It essentially automates the process "word-of-mouth" recommendations: items are recommended to a user based upon values assigned by other people with similar taste. The system determines which users have similar taste via standard formulas for computing statistical correlations. Social Information filtering overcomes some of the limitations of content-based filtering. Items being filtered need not be amenable to parsing by a computer. Furthermore, the system may recommend items to the user which are very different (content-wise) from what the user has indicated liking before. Finally, recommendations are based on the quality of items, rather than more objective properties of the items themselves [36, 38]. Some of the most popular systems using content-based filtering are WebWatcher [30], and client-side agent Letizia [3].

Rule-based filtering. The users are asked to answer a set of questions. These questions are derived from a decision tree, so as the user proceeds to answer them. What he finally receives is a result (e.g. list of products) tailored to his needs. Content-based, rule-based, and collaborative filtering may also be used in combination, for deducing more accurate conclusions. Some of the rule-based filtering drawbacks are: User input may be subjective and prone to bias, explicit (and non-binary) user ratings may not be available, profiles may be static and can become outdated quickly, and for large systems it becomes burdensome to manage. Related interesting systems include Dell, Apple Computer, Amazon.com, CDNOW, and Broadvision [38, 36, 10, 26].

Collaborative filtering. Systems invite users to rate the objects or divulge their preferences and interests and then return information that is predicted to be of interest to them. This is based on the assumption that users with similar behaviour (e.g. users that are rate similar objects) have analogous interests.There are two general classes of collaborative filtering algorithms, memory-based methods and model-based methods [10, 26, 34, 35]. Moreover, the goals in a collaborative filtering system are basically focused upon the reduction of computation time, the increase of the extent in which predictions can be computed in parallel, and the increase of prediction accuracy. Collaborative filtering can further refine the process of giving each individual personal recommendation compared to rule-based filtering. It overcomes the drawbacks of the content-based filtering because it typically does not use the actual content of the items for recommendation. It usually works based on assumptions. With this algorithm the similarity between the users is evaluated based on their ratings of products, and the recommendation is generated considering the items visited by nearest neighbors of the user. In its original form, the nearest-neighbor algorithm uses a two-dimensional user-item matrix to represent the user profiles. This original form suffers from three problems, *scalability*, *sparsity*, and *synonymy* [37, 40]. Some more highlighted drawbacks of collaborative filtering are focused upon: (a) Collaborative-filtering techniques are often based in matching in real-time the current user's profile against similar records obtained by the systems over time from other users. However, as noted in recent studies, it becomes hard to scale collaborative filtering techniques to a large number of items, while maintaining reasonable prediction performance and accuracy. Part of this is due to the increasing sparsity in the data as the number of items increase. One potential solution to this problem is to first cluster user records with similar characteristics, and focus the search for nearest neighbours only in the matching clusters. In the context of Web personalization this task involves clustering user transactions identified in the preprocessing stage; (b) traditional collaborative filtering does little or no offline computation, and its online computation scales with the number of customers and catalog items. The algorithm is impractical on large data sets, unless it uses dimensionality reduction, sampling, or partitioning – all of which reduce recommendation quality; (c) user input may be subjective and prone to bias; (d) explicit (and non-binary) user ratings may not be available; (e) profiles may be static and can become outdated quickly; (f) they are not able to recommend new items that have not already been rated by other users. An object will become available for recommendation only when many users have seen it and rated it, making it part of their profiles first ("latency problem"); (g) they are not satisfactory when dealing with a user that is not similar enough with any of the existing users [27, 38, 41, 39]. Some systems applied with this technique are Yahoo, Excite, Microsoft Network, Net Perceptions [26, 27].

Web-usage Mining. The typical sub-categorization of the Web mining research field falls into the following three categories: Web-content mining, Web-structure mining, and Web usage mining. The prerequisite step to all of the techniques for providing users with recommendations is the identification of a set of user sessions from the raw usage data provided by the Web server. Web usage mining is the only category related to Web Personalization. This process relies on the application of statistical and data mining methods to the Web log data, resulting in a set of useful patterns that indicate users' navigational behavior. The data mining methods that are employed are: Association rule-mining, sequential pattern discovery, clustering, and classification. Given the site map structure and usage logs, a Web usage miner provides results regarding usage patterns, user behaviour, session and user clusters, clickstream information, and so on. Additional information about the individual users can be obtained by the user profiles [42, 26, 5]. The overall process can be divided into two components. (a) The *offline component* is comprised of the preprocessing and data preparation tasks, including data cleaning,

filtering, and transaction identification, resulting in a user transaction file, and (b) the *data mining stage* in which usage patterns are discovered via specific usage mining techniques such as association-rule mining, association-rule discovery and usage clustering [41]. The increasing focus on Web-usage mining as the time passes derives from some key characteristics which are summarized as follows: (a) the profiles are dynamically obtained, from user patterns, and thus the system performance does not degrade over time as the profiles age; (b) using content similarly alone as a way to obtain aggregate profiles may result in missing important relationships among Web objects based on their usage. Thus, Web usage mining will reduce the need for obtaining subjective user ratings or registration-based personal preferences; (c) profiles are based on objective information (how users actually use the site); (d) there is no explicit user ratings or interaction with users (saves time and other complications); (e) it helps preserve user privacy, by making effective use of anonymous data; (f) the usage data captures relationships missed by content-based approaches; (g) it can help enhance the effectiveness of collaborative or content-based filtering techniques. Nevertheless, usage-based personalization can be problematic when little usage data is available pertaining to some objects or when the site content attributes of a site must be integrated into a Web mining framework and used by the recommendation engine in a uniform manner [38, 27]. Noteworthy applications are Alta-Vista, Lycos, WebSift, and SpeedTracer [43, 26].

Demographic-based filtering. This specific technique could be roughly described as an approach that uses demographic information to identify the types of users that prefers a certain object and to identify one of the several pre-existing clusters to which a user belongs and to tailor recommendations based on information about others in this cluster [34, 35].

Agent technologies. Agents are processes with the aim of performing tasks for their users, usually with autonomy, playing the role of personal assistants [44, 2]. Agents usually solve common problems users experience on the Web such as personal history, shortcuts, page watching and traffic lights [4]. Some of the agents' main characteristics could be distinguished according to their abilities used and according to the tasks they execute. The former include characteristics such as *intelligence*, *autonomy*, *social capacity* (inter-agent communication), and *mobility*; while the latter classify the agents into *information filtering agents*, *information retrieval agents*, *recommendation agents*, *agents for electronic market*, and *agents for network management* [44].

Since the mobility dimension is also incorporated in this paper and therefore addresses vital needs as to locate the required information, on time, under any circumstances the use of intelligent mobile agents for the a given wireless environment could be proved ideal for implementing various Web personalization processes. Intelligent mobile agents are identified by some specific capabilities focused upon: (a) *Reduction of the network load*, instead of relying on numerous communication protocols to achieve network interaction, which would increase the network traffic, mobile agents can carry with them the data that is required for an interaction and process it locally; (b) o*vercoming network latency*, mobile agents can help in critical real-time systems where a response to environment changes is required in real time and latencies will not be tolerated. Mobile agents can be dispatched from a central controller to act locally and directly execute the controller's directions; (c) *asynchronous and autonomous execution*, after a task is assigned to a mobile agent, the agent will be dispatched into the network and become independent of the creating process. It can operate asynchronously and autonomously, relieving its owner from having continuously an eye on its activities. The agent's owner will be able to collect it at some later time, if needed; and (d) *dynamic adaptation*, mobile agents are capable of monitoring the environment in which they operate and react to the changes accordingly. Last but not least, (e) mobile agents are naturally heterogeneous, robust and fault-tolerant, and able to encapsulate protocols considered vital for the universal development of open,

modular, ubiquitous and personalized mobile learning adaptive hypermedia applications [39, 5, 2, 17]. Pioneer personalization systems implemented with intelligent agents are: ARCHIMIDES, Proteus, WBI, BASAR, 1:1 Pro, Haystack, eRACE, mPersona, Fenix system, and SmartClient [45, 2, 44]

Cluster Models. These types of techniques are found mostly in the area of eCommerce and could be characterized as eCommerce recommendation algorithms. To find customers who are similar to the user, cluster models divide the customer base into many segments and treat the task as a classification problem. The algorithm's goal is to assign the user to the segment containing the most similar customers. It then uses the purchases and ratings of the customers in the segment to generate recommendations. The segments typically are created using a clustering or other unsupervised learning algorithm, although some applications use manually determined segments. Using a similarity metric, a clustering algorithm groups the most similar customers together to form clusters or segments. 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 customer each. They then repeatedly match customers to the existing segments, usually with some provision for creating new or merging existing segments. For very large data sets – especially those with high dimensionality – sampling or dimensionality reduction is also necessary. Once the algorithm generates the segments, it computes the user's 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 [30]. Cluster models have better online scalability and performance than collaborative filtering because they compare the user to a controlled number of segments rather than the entire customer base. The complex and expensive clustering computation is run offline. However, recommendation quality is relatively poor. To improve it, it is possible to increase the number of segments, but this makes the online user segment classification expensive. Typical examples of eCommerce systems are Amazon.com [29], Dell [26], and IBM.com [28].

6. Conclusion

With the emergence of mobility the user requirements have been altered significantly introducing new personalization approaches and methodologies, satisfying interactive access on demand from an “anywhere, anytime and anyhow” basis.

The main scope of this paper was to make an up-to-date analysis of Web personalization insights and technological paradigms regarding both the desktop and mobile user and to provide a basis for the consideration and employment of research models coming from different research areas, such as social and cognitive psychology, attention economy, and human computer interaction, that could incorporate other dimensions to the meaning of Web personalization resulting to a sustainable and concrete definition. A review of user profiling, a main personalization instrument, the five Web personalization categories (Link personalization, Content personalization, Context personalization, Authorized personalization, and Humanized personalization) and the Web personalization paradigms, namely content-based filtering, rule-based filtering, collaborative filtering, Web-usage mining, demographic-based filtering, agent technologies (applicable mostly in mobile and wireless applications), and cluster models (applicable mostly in eCommerce systems) have been analyzed, compared and contrasted.

References

[1] Mulvenna M. D., Anand S. S., & Buchner A. G. (2000). *Personalization on the net 23using web mining*, Communications of the ACM, 43, 8 (August), 123–125.

[2] Panayiotou C., & Samaras G, (2004), *mPersona: Personalized Portals for the Wireless User: An Agent Approach*, Journal of ACM/ Baltzer Mobile Networking and Applications (MONET), special issue on "Mobile and Pervasive Commerce".

[3] Chaffee J., & Gauch S. (2000). *Personal Ontologies for Web Navigation*, ACM 2000, 1-58113-320-0/00/11.

[4] Barett R., Maglio P., & Kellem D. *How to personalize the Web*, [online], http://www.raleigh.ibm.com/wbi/wbisoft.htm.

[5] Cingil I., Dogac A., & Azgin A. (2000). *A broader approach to personalization*, Communications of the ACM, Vol. 43, No. 8.

[6] Blom J. (2000). *Personalization – A Taxonomy*, ACM 2000. ISBN:1-58113-248-4.

[7] Kim H. *Web Personalization*, critical survey paper, [online], http://my.fit.edu /~hokim/file/depthpaper.pdf.

[8] Kim W. (2002). *Personalization: Definition, Status, and Challenges Ahead*, Published by ETH Zurich, Chair of Software Engineering JOT, ©2002, Vol. 1, No. 1.

[9] Ricci C. (2004). *Personalization is not a technology: Using Web personalization to promote your business goal*, [online], http://www.usabilityviews.com/uv006523.html.

[10] Wang J., & Lin J. (2002). *Are personalization systems really personal? – Effects of conformity in reducing information overload*, Proceedings of the 36th Hawaii International Conference on Systems Sciences (HICSS'03), 0-7695-1874-5/03, © 2002 IEEE.

[11] De Bra P., Aroyo L., & Chepegin V. (2004). *The next big thing: Adaptive Web-based systems*, Journal of Digital Information, Vol. 5 Issue 1, Article no. 247.

[12] Brusilovsky P., Nejdl W. (2004). *Adaptive Hypermedia and Adaptive Web*, © 2004 CSC Press LLC.

[13] Conlan O., Lewis D., Higel S., O'Sullivan D., & Wade V. *Applying Adaptive Hypermedia Techniques to Semantic Web Services Composition*, [online], http://wwwis.win.tue.nl/ah2003/proceedings/www-5/.

[14] Barna P., Frasincar F., Houben G., & Vdovjak R. *Methodologies for Web Information System Design*, [online], http://citeseer.ist.psu.edu/680105.html.

[15] Billsus D., Brunk C., Evans C., Galdish B., & Pazzani M. (2002). *Adaptive interfaces for ubiquitous Web access*, Communications of the ACM, Vol. 45, No. 5.

[16] Wu J., His E., Kate W., & Chen P. (2000). *A framework for Web content adaptation*, [online], http://www.w3.org/2000/10/DIAWorkshop/wu.htm.

[17] Billsus D., Pazzani M., & Chen J. *A learning agent for wireless news access*, [online], http://www.fxpal.com/people/billsus/pubs/iuinews.pdf.

[18] Anderson C., Domingos P., & Weld D. (2001). *Personalizing Web sites for mobile users*, [online], http://citeseer.ist.psu.edu/444187.html.

[19] Lankhorst M.M., Kranenburg, Salden A., & Peddemors A.J.H. (2002). *Enabling Technology for Personalizing Mobile Services*, Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35'02).

[20] O'Connor A., Wade V., & Conlan O. *Context-Informed Adaptive Hypermedia*, [online], http://pace.dstc.edu.au/cw2004/Paper9.pdf.

[21] Volokh E. (2000). *Personalization and Privacy*, from the Communications of the Association for Computing Machinery, vol. 43, issue 8, at 84.

[22] Korkea-aho M. (2000). *Context-aware applications survey*, [online], http://www.hut.fi/~mkorkeaa/doc/context-aware.html.

[23] Anderson C., Domingos P., & Weld D. (2001). *Adaptive Web navigation for wireless devices*, [online], http://www.cs.washington.edu/ai/proteus/ijcai01.pdf.

[24] Adomavicious G., & Tuzhilin A. (1999). *User Profiling in Personalization Applications through Rule Discovery and Validation*, ©ACM 1999 1-58113-143-7/99/08.

[25] Kotinurmi P. *User Profiles and their Management*, [online], http://www.tml.hut.fi/ Studies/Tik-111.590/2001s/papers/paavo_kotinurmi.pdf.

[26] Eirinaki M., & Vazirgiannis M. (2003). *Web Mining for Web Personalization*, ACM Transactions on Internet Technology, Vol. 3, No. 1, Pages 1-27.

[27] Mobasher B., Dai H., Luo T., Sun Y., & Zhu J. (2000). *Combining Web usage and Content mining for more effective personalization*, [online], http://citeseer.ist.psu.edu/ mobasher00combining.html.

[28] Karat M.C., Brodie C., Karat J., Vergo J., & Alpert S.R. (2003). *Personalizing the user experience on ibm.com*, IBM Systems Journal , Vol. 42, No. 4.

[29] Rossi G., Schwade D., & Guimaraes M.R. (2001). *Designing Personalized Web Applications*, ACM 1-58113-348-0/01/0005.

[30] Perkowitz M, & Etzioni O. (2003). *Adaptive Websites*, Towards Adaptive Web sites: Conceptual Framework and case study, [online], http://www.cse.buffalo.edu/ ~sbraynov/ seninar2003/presenttations/adaptive.pdf.

[31] Dagger D., Wade V., & Conlan O. (2003). *Towards "anytime, anywhere" learning: The role and realization of dynamic terminal personalization and adaptive eLearning*, [online], http://www.cs.tcd.ie/Owen.Conlan/publications/EdMedia2003v1_ Conlan.pdf.

[32] Hjelsvold R., Vdaygiri S., & Leaute Y. (2001). *Web –based personalization and management of interactive video*, ACM 1-58113-348-0/01/0005.

[33] Kaplan B., Farzanfar R., & Freeman H.R. *Research and ethical issues arising from ethnographic interviews of patients' reactions to an intelligent interactive telephone health behavior advisor*, [online], http://is.lse.ac.uk/Support/ifip_wg82/ StLouis/kaplan.pdf.

[34] Pazzani J. M. (2005). *A framework for collaborative, content-based and demographic filtering*, [online], http://www.ics.uci.edu/~pazzani/Publications/ AIREVIEW.pdf.

[35] Basilico J., & Hofmann T. (2004). *Unifying collaborative and content-based filtering*, Proceedings of the 21st International Conference on Machine Learning, Banff, Canada.

[36] Sharbanand U., & Maes P. (1998). *Social information filtering: Algorithms for automating "Word of Mouth"*, [online], http://citeseer.ist.psu.edu/ shardanand95social.html.

[37] Shahabi C., & Chen Y. (2003). *Web information personalization: Challenges and approaches*, [online], http://infolab.usc.edu/DocsDemos/DNIS2003.pdf.

[38] Mobasher B., Dai H., Luo T., Nakagawa M., & Wiltshire J. (2000). *Discovery of aggregate usage profiles for Web personalization*, [online], http://maya.cs.depaul.edu/ ~mobasher/papers/webkdd2000.pdf.

[39] Vozalis E., Nicolaou A., & Margaritis K. G. (2001). *Intelligent Techniques for Web Applications: Review and Educational Application*, presented at the Fifth Hellenic-European Conference on Computer Mathematics and its Applications (HERCMA), Athens, Greece.

[40] Papagelis M., Plexousakis D., Rousidis D., & Theoharopoulos E. (2004). *Qualitative Analysis of User-based and Item-based Prediction Algorithms for Recommendation Systems*, CIA 2004: 152-166.

[41] Mobasher B., Colley R., & Strivastava J. (1997). *Automatic Personalization Based on Web Usage Mining*, [online], http://maya.cs.depaul.edu/~mobasher/personalization.

[42] Deshpande M., & Karypis G. (2004). *Selective Markov models for predicting Web page accesses*, ACM Transactions on Internet Technologies, Vol. 4, No. 2, Pages 163-184.

[43] Pierakkos D., Paliouras G., Papatheodorou C., & Spyropoulos D.C. (2001). *KOINOTITES: A web usage mining tool for personalization*, [online], http://iit.demokritos.gr/~paliourg/papers/PC-HCI2001.pdf.

[44] Delicato F., Pirmez L., & Carmo L. (2001). *Fenix – personalized information filtering system for WWW pages*, Internet Research: Electronic Networking Applications and Policy, Vol. 11, No. 1, pp. 42-48.

[45] Pu P., & Faltings B. (2002). *Personalized Navigation of Heterogeneous Products Spaces using SmartClient*, © 2002 ACM 1-58113-459-2/02/0001/.