Personalized Information Systems: Enabling Technologies and Architecture

José Luís Reis
Centro Algoritmi
Universidade do Minho
Guimarães - Portugal
jlreis@cvrvv.pt

João Álvaro Carvalho
Centro Algoritmi and Departamento de Sistemas de Informação
Universidade do Minho
Guimarães - Portugal
jac@dsi.uminho.pt

Abstract—**Personalized Information Systems are ICT applications that exhibit personalized behaviour, adjusted to the preferences or needs of their users. Personalized Information Systems can bring up business benefits related to customer relationship or to increased efficiency of organizational work. This work characterizes the Personalized Information Systems in what concerns types of personalized features, presents the enabling technologies, and suggests a general architecture for Personalized Information Systems.**

Keywords- personalization, information systems, technologies, architecture.

I. INTRODUCTION

With the globalization of Information and Communication Technologies (ICT) (due mainly to the widespread of Internet/Web based services), personalization became a reality in many Web sites and in some organizational ICT applications. The personalization of applications is a complex process and this work aims to address several issues related with types of personalization, enabling technologies required to achieve personalization and a general architecture for personalized information system.

II. PERSONALIZATION

There are several areas of activity where addresses personalization, in academia or in organizations, for example: management, economics, marketing and sales, political science, sociology, anthropology, medical sciences, clinical information systems and computer science [1], [2]. The main goal of personalization in ICT applications is to address the specific and differentiated needs and preferences of the persons that interact with them.

Personalized ICT applications (also named as Personalized Applications – PA) facilitate the interaction among the various actors that cooperate within an organization, and the interaction with information providers and information recipients. It allows people to have easier access to content and services.

The "individuality" of the consumer information is the key aspect in the personalization definition, implying that individual attributes, such as identity, preferences, constraints and provision (example, location and presence) have a potential impact on the personalization process [3]. The personalization is done by the intelligence that is at the center of the communication.

III. PERSONALIZATION TYPOLOGIES

According to the personalized features and purpose of a PA it is possible to establish different typologies to achieve personalization based on different aspects, including [1], [4]:

- What is personalized? The system characteristics that will be personalized (for example: content, user interface, features and channels of information delivery).

- Who are the personalization targets? Individuals or groups/categories of specific individuals.

- Who makes the personalization (user or system, or both)? **Implicit or adaptive/proactive**, if the personalization is done automatically by the system when, for example the system with a collaborative filtering techniques discover that if a user likes music from Madonna also likes music from Beyoncé, because he have the some profile characteristics from other user that likes de same. **Explicit or reactive** if the user participates in providing information or making choices that will allow the system to adapt, for example if the user fills an electronic form with the type preferences in a book shop (poetry, romance, drama, etc.).

- How is personalization achieved? Whether for an individual user or a group of users, there are two possible ways to get to personalization: a) through processes of personalization that not require intervention from users, called **observational** [5] **or implicit** personalization, for example the system presents a recommendation list shopping based in previous purchases; b) personalization processes that somehow require the participation of users. If the user is interrupted in the work processes that are involved to provide answers to questions that are placed upon it, it's called **intrusive or explicit** personalization [6].

IV. TECHNOLOGIES FOR PERSONALIZATION

Personalization of ICT applications rely on a profile of their users. The profile covers relevant aspects of users, such as identification, characteristics, preferences, interest and typical behavior [7]. To build and explore this profile, several different approaches and technologies can be used, adequate to the different sources and to the actions to be taken. The storage of this profile is another issue in a PA.

The main technologies for personalization include [4]: **collaborating filtering** enable to create recommendations based on implicit or explicit collaboration between users and

the system, collaborating filtering is based on the principles that relate clients that might have similar profiles; **rules-based filtering** generate profiles based in users behavior and enabled the system to present items to groups of individuals according to previously defined standards behavior for these groups; **content-based filtering** allows to make recommendations of items that have properties similar to other items that have been selected by the users in the past; **clustering techniques** allow to create associations between users, that are obtained from similar transactions or from the number of times certain situations occurred in transactions.

V. INFORMATION GATHERING FOR PERSONALIZATION

Information to build an user's profile is gathered from interaction between the user and the PA. Interacting events include [8]:

- Display of requested web pages.

- Search queries entered and answers to surveys.

- Videos and music downloaded or played.

- Responses to advertisements displayed. Web ads in network environments from advertising media companies (example: "text ads" from Google's AdSense or AOL's Advertising.com).

Each time one of these events occurs, there is a relationship established between the user and the PA (and the PA's support server) and with the advertising services that the PAS is cooperating with. Interaction events involve mouse clicking, keyboard typing and screen touching [9]. A record of these actions (including what has been typed or pointed) is kept in log records. This information used then to build the users' profile. Personalization can address individuals or groups of users who share the some identified characteristics or preferences.

The following paragraphs present the main information sources for personalization.

Clickstream analysis - Clickstream is a record of typing's and pointing/selections made by a user while interacting with an ICT application. The clickstream provides information on which pages were selected, what information has been obtained, the time (data/hour) pages were selected, the time spent on each page, or even the interests associated with the types of stores (directed purchases, purchases from research, browsing for pleasure and knowledge construction) [10] and selected products on the Web. User's interests are deduced from addresses obtained by the clickstream analysis.

Cookies - Internet cookies are sets of text that store detailed information about users. Although they do not identify individuals, cookies provide information that combines a computer, a user and a browser [11]. The cookie is usually created, combined and sent by a Web server to the users' computers when they visit their sites. It contains information about the visitor, detailing what they did when they navigated the site, such as: date of visit, pages visited, page views over time, etc. [12], [13]. When visitors return to the site that placed the cookie (using the computer where the cookie is located), the system accesses to the information contained in the cookie, and acts as it remembers the last interaction [14] .

Electronic mail (e-mailing) - The most used Internet application is the electronic mail (e-mail). It is used for sending and reading massages by around 94% of Internet users [15]. Some companies use expert systems for processing information in e-mail messages for marketing purposes (for example www.constantcontact.com or www.e-goi.com). Examples of information that can be obtained this way include: who opens the emails, who do forwards, which links are clicked, who joined a particular campaign, conversion rates, etc. After being processed, this information can be used to update a users profile as well to trigger marketing actions [16].

Electronic forms - Electronic forms (e-forms) are used by organizations to collect information about a particular individual or organization with whom a certain relationship is established. The information obtained about customers though the electronic forms enable to improve efficiency and services associated with the interaction with users. Improvements might address aspects such as reducing wait times, reducing transaction costs, elimination of support paper, printing, processing, distribution and storage, reduce the negative impact environment and can ensure that data are collected in a more accurate and complete with a significant cost savings and time [17]. Electronic forms can be activated when a user registers on a site to perform a transaction (electronics purchases, join a society, etc.), register for a specific event (, show, conference, training activity, etc.). The data collected through electronic forms allow determine users' preferences. The information obtained is then used to perform targeted marketing actions [18].

Transaction records – Transactions records are records created and kept by transaction systems (TPS - Transaction Processing Systems) [19]. These records can be used to build the profiles of target customers. Techniques such as RFM (Recency, Frequency and Monetary) can be used to define the potential customers or users in a given period of time, on the dates of the last transaction, the transactions frequency and the transactions value [20], [13], which may be related to the price. There are several models based on mathematics and statistics approaches that use historical data to predict the buying behavior of customers. Transaction records provide information about aspects such as dates of purchase, items purchased, prices, quantities purchased, payment methods, discounts, fees, grants, who sold, who bought, the place of purchase, etc. The majority of transactions are from support services to customers, orders, billing and accounting data [21] and its treatment is essential to the managerial activities.

Demographic, psychographic and geo-referenced information - Demographic and psychographic information can be obtained from electronic forms, from scanned documents or from transaction records. Geo-referenced information can be obtained from topographic maps and from GPS technology (Global Positioning Systems). The information associated with a user profile can be more or less dynamic, depending on their characteristics. The

characteristics of the demographic information (name, address, gender, age, salary, number of identity card, tax identification number, etc.) [22], differentiate users for responses to messages that are sent, the same is true for the history of their behaviour, as if their histories are different emotional behaviour, then also respond differently to messages [23], hence the importance of data psychographics (lifestyles, values, interests, social class, etc.) that must be combined with the population to obtain the profile of individual users. With the popularization of geo-referenced technologies (mobile phones with GPS, geo-referenced maps on the Web, vehicles with GPS systems, etc.), also became important to determine the location of users.

Biometrics - So far the best way to recognize in an absolutely unequivocal one person is through biometrics [24], [25]. It is estimated that by 2020, voice recognition and touch-screen interfaces will be more prevalent and accepted in the Internet [26]. There are various technologies associated to biometric authentication. They include control by the eye tracking (iris/retina), facial recognition, fingerprint recognition, voice recognition, recognition of the lips, the recognition of hand geometry, recognition through a DNA sample. More recent advancements in this area include interfaces sensitive to the emotions (Emotionally Sensitive Interfaces - ESI) and recognition by the ears [27]. These different types of biometrics can be combined among them (fusion software) to provide more reliable identifications and authentications [28]. Authentication thought biometric technologies, although more invasive, will redefine the personalization technologies [29], [26].

VI. METHODS AND TECHNOLOGIES FOR PERSONALIZATION

When the user profile is available, it is necessary to analyze the information associated with this profile in order to provide interfaces that enabled recommend documents, content, shopping, or specific actions to that user [30] (Chiu, 2001).

The main personalization techniques categories are associated with requests and queries from users, tailored interfaces, and personalized information.

The main technologies used for personalization are based on filtering and rules-based technology [31], [30], because, as shown by the analysis of the information presented below, on personalization technologies the management of user preferences and personalized interfaces use this type of technology [32].

Behavior monitoring - Provide mechanisms for identifying users and monitoring their behavior. Technologies: clickstream tracking, user recognition, pattern recognition (for example, case-based reasoning).

Models of preferences – Personalized Data Bases - Interactive systems to record and manage data in electronic. They required intensive computation. Determining customer value for different attributes and make alternate combinations that enable defining what are the goods or services to be delivered to users. Technologies: Collaborative filtering, filtering based on content (content-based), Web hosting (Web

housing), intelligent agents, data mining, profiling based on rules (rules-based), statistical analysis traffic, recommendations systems and hierarchies of Bayes.

Personalized interfaces - Include user interfaces design, digital design oriented to the human being and adaptive hypermedia. Technologies: content management, audio and video streams (streaming), user information, filtering, design of interfaces according to preferences (Web parts) and personalized search.

Prediction models - Enabled predicting customer response to targeted offers, as well as the likelihood of a user to click an advertisement. Determine the effect of prices and promotions on the probability of making a purchase. Time prevision associated with the permanence of a user in a particular Web screen/page. Technologies: predictive models, regression analysis, neural networks and Bayesian networks.

Customer support - Personalize applications in real time (just-in-time) to support e-commerce, which capitalize the knowledge of the customer's location, its business and the environment that surrounds it. Technologies: real-time customer support, wireless services to clients and software agents.

Models of grouping/segmentation and classification of customer's - Allow defining the characteristics of a segment and the segment belong to the users. Enable organizations to classify customers. Technologies: grouping (clustering), classification algorithms and regression trees.

VII. A GENERIC ARCHITECTURE FOR PERSONALIZED INFORMATION SYSTEMS

The PA architecture represents the main components of a PA and the most important interactions among them.

The components enabled the personalization features and assure the execution of the personalization process as described in the previous section. The personalization of a system always involves implementation of a support service that optimizes the user interaction with the information system [33].

The proposed architecture presents a set of components of various types (memory, processing technologies and interfaces). They improve a better communication between the user and the system and they are represented by different symbols in figure 1.
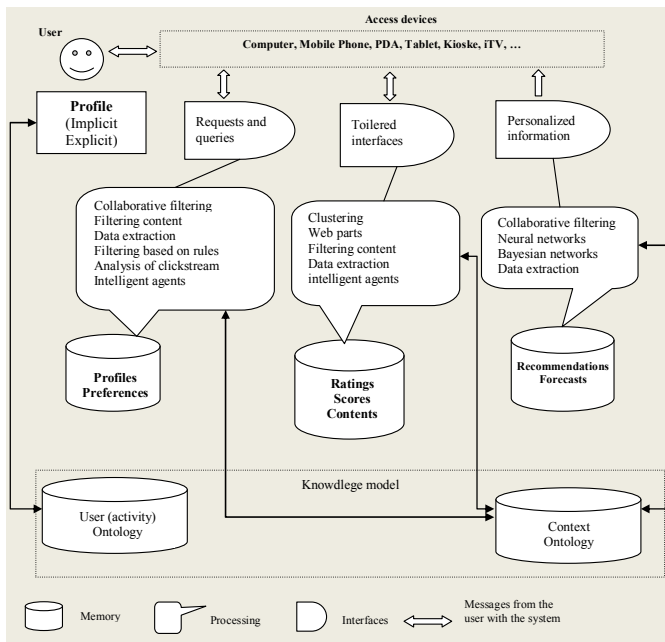
Figure 1 – Personalized information system architecture

As can be seen in figure 1, the techniques categories of personalization represent the interface components in the architecture (requests and queries from users, tailored interfaces and personalized information), and their cross-processing with technologies that transforms the data and recorded it in the memory components (profiles, preferences, ratings, scores, content, recommendations and forecasts) [22], [34].

Users access the system through different devices (for example: personal computers, mobile phones, PDAs, tablets, multimedia kiosks, interactive TV, etc.). Users interact with the system by posing queries and requests for information. Different technologies can be used to adapt to different interfacing devices and to present information tailored to user's profile. Users profiles and preferences are recorded in the memory after having been processing by the profile building components (that make use of different technologies that can be used separately or simultaneously), including collaborative filtering, content filtering, data extraction, rules-based filtering, clickstream analysis and intelligent agents.

A personalized system should allow interfaces adapted to users. The system must contain custom adaptive features, such as adapting the selection options in menus according to user's profile (for example, through quick access buttons), interfaces with visual components adapted to users preferences to reduce the cognitive load during a Web search, presenting recommendations lists that fit users preferences and needs. To construct the adapted interfaces different technologies are used. They including clustering, Web parts, content-filtering, data mining and intelligent agents that transform the data stored with the ratings, scores and content that allow the personalization.

A personalized system must provide personalized information to the user needs. The system must have the capacity to process clues obtained from the user activities (for example, the flow of clicks or transactions) through processing techniques based on mathematics and statistics models, that rely on different technologies (collaborative filtering, neural networks, Bayesian networks, data mining, etc.).

Technologies such as ontologies are normally used in a PA to improve the knowledge about the users; through context ontology (for example, the museums tourism times, the types of museums, the hours of guided tours, etc.) and user ontology (preferences about the types of museums, languages spoken, etc.) [4]. Technologies can be used in isolation or in combination with other technologies. Figure 1 attempt to illustrate this aspect. Although the architecture does not include the process model, in a PA the obtained information (profiles, rules, ratings, content, activities, scores, etc.) must be recorded in order to create knowledge bases based on data patterns and users behaviors to get better interaction with those users, providing personalized information, including tailored interfaces that may contain recommendations about users interests, for example, restaurants, places to visit, lists, etc. [34], [35], [36].

VIII. CONCLUSIONS AND FUTURE WORK

This article presents a synopsis of technologies relevant to personalization. This has been done by collecting information from different sources and presenting them in a consolidated way. This dispersion of information about personalization technologies results from non-integrated perspective existing in studies about personalization. This study sought to fill this gap, producing a systematic framework that integrates various levels of intervention that are relevant to personalization.

The rate of development of information technology with regard to hardware components, interface devices and new ways of processing is very high. The consequence of this evolution is the rapid obsolescence of technologies. Despite some stability of the proposed personalized information system architecture, the enabling technologies are subject to change resulting from technological innovations.

Several new paths to personalized systems can be envisaged, such as: software support for mobile devices; processing individual information (individualization of information); new forms of user authentication integrated into multiple sites; the possibility of transactional information as a result of user experiences, might be crossed (with the authorization of their own) by various organizations, force companies to turn to sources of information from the cookies, clickstream analysis, transaction information, emails, e-forms, etc. which will add demographic and psychographic information vital to the organizational marketing communications.

The framework established in this work provides a basis for reviewing and assessing existing technologies. Moreover, decisions related to personalization may refer to the following areas of future exploration: to achieve the objectives of the personalization of systems, organizations will have to make investments in the use of techniques and tools such as definition of the user profiling, collaborating filtering, grouping (clustering), the e-mail management tools and data

mining, just so the management of the business can become more intelligent.

REFERENCES

[1] Fan, H., & Poole, M. S. (2006). What Is Personalization? Perspectives on the Design and Implementation of Personalization in Information Systems. *16*, 179-202.

[2] Mittman, B., Borman, L., & Hayes, W.-B. &. (1975). *Personalized Data Base Systems*. Los Angeles: Melville Publishing Company.

[3] Toth, K., & Nagboth, S. R. (2002). *Constraint-Based Personalization Model: Multi-Channel Messaging*. Oregon, USA: Dept. of Computer Science, Oregon State University.

[4] Anand, S. S., & Mobasher, B. (2005). Intelligent Techniques for Web Personalization.

[5] Mulvenna M., A. S. (2000). Personalization on the Net using Web Mining. *43(8)*, 122–125. Communications of the ACM.

[6] Perugini, S., & Ramakrishnan, N. (2002). *Personalizing Interactions with Information Systems*. Blacksburg - Virginia: Department of Computer Science.

[7] Hellmund, M. (2003). Smart Personalization for Wireless Applications.

[8] Story, L. (2008). *How Do They Track You? Let Us Count the Ways*. Retrieved on 12. 02.2011, from The New Yourk Times Business Day Technology: http://bits.blogs.nytimes.com/2008/03/09/how-do-they-track-you-let-us-count-the-ways/

[9] Micarelli, A., Gsparetti, F., Sciarrone, F., & Gauch, S. (2007). Personalized Search on the World Wide Web. Springer Berlin / Heidelberg.

[10] Moe, W. (2003). Buying, Searching, or Browsing: Differentiating Between Online Shoppers. Using In-Store Navigational Clickstream.

[11] Google. (2008). *Cookies & Google Analytics*. Retrieved on 03.01.2011, from Google Analytics: http://code.google.com/apis/analytics/docs/concepts/gaConceptsCookies.html#cookiesSet

[12] Rud, O. (2001). *Data Mining Cookbook*. Wiley.

[13] Dionísio, P., Rodrigues, V., Faria, H., Canhoto, R., & Nunes, R. (2009). *b-Mercator - Blended Marketing* . Dom Quixote.

[14] Markov, Z., & Larose, D. (2007). Data Mining The Web - Uncover Patterns in Web Contente, Structure, and Usage. John Wiley & Sons.

[15] Project, P. I. (2010). *Trend Data* . Retrieved on 20 de 03 from 2011, de Pew Internet: http://www.pewinternet.org/Trend-Data/Online-Activites-Total.aspx

[16] Kruger, J., & Epley, N. (2005). Egocentrism Over E-Mail: Can We Communicate as Well as We Think? *Vol. 89, No. 6*, 925–936.

[17] IBM. (2007). *IBM Forms - electronic forms*. Retrieved on 16.03.2011, from ftp://ftp.software.ibm.com/pub/lotusweb/forms/LotusForms-RecessionProof_final.pdf

[18] Choi, S.-Y., Stahl, D. O., & Whinston, A. B. (22.22.2008 de 2003). *The Economics of Electronic Commerce*. Macmillam Technical Publishing.

[19] Laudon, K., & Laudon, J. (2006). *Management Information Systems - Managing the digital firm* (Ninth ed.). Pearson Prentice Hall.

[20] Hughes, A. (2000). *Strategic DataBase Marketing* (2ª Ed ed.). McGraw-Hill.

[21] Blattberg, R. C., Kim, B.-D., & Neslin, S. A. (2008). *Database Marketing, Analyzing and Managing Customers*. Springer.

[22] Jrad, Z., Aufaure, M.-A., & Hadjouni, M. (2007). A Contextual user model for Web personalization.

[23] Moon, Y. (2002). Personalization and Personality: Some Effects of Customizing Message Style Based on Consumer Personality.

[24] Cutting, D. (2006). *Identity Management in Context-Aware Intelligent Environments* . University of Sydney, School of Information Technology.

[25] Murthi, B. P., & Sarkar, S. (2003). The Role of the Management Sciences in Research on Personalization. *Vol. 49, No. 10*, 1344–1362. (INFORMS, Ed.) The University of Texas at Dallas.

[26] Project, T. P. (2008). *The Future of the Internet III*. Retrieved on 20 de 03 de 2011, de The Pew Internet & American Life Project: http://www.pewinternet.org/~/media/Files/Reports/2008/PIP_FutureInternet3.pdf.pdf

[27] Mims, C. (2010). *The Hot New Thing in Biometric Security is... Ears*. Retrieved on 20.03.2011, from Technology Review: http://www.technologyreview.com/blog/mimssbits/25787/

[28] Jonietz, E. (2004). *Boosting Biometrics*. (MIT) Retrieved on 20.03.2011, from Technology Review: http://www.technologyreview.com/computing/13618/?mod=related

[29] Chia, C., & Garcia, J. (2002). *The personalization challenge in public libraries: perspectives and prospects*. Retrieved on 10.11.2008, from PLIN - Public Libraries International Network: www.public-libraries.net/html/x_media/pdf/personalisation_engl.pdf

[30] Chiu, W. (17 de 04 de 2001). *Web site personalization*. (IBM, Editor) Retrieved on 30.09.2010, from IBM - Technical library view: http://www.ibm.com/developerworks/websphere/library/techarticles/hipods/personalize.html

[31] Datta, A., Thomas, H., & VanderMeer, D. (2001). Personalizing Your Web Site - a How-to Guide. The Penn State eBusiness Research Center.

[32] Amoroso, D., & Reinig, B. (2003). Personalization management systems. Big Island - Hawaii.

[33] Thomson, L. (2005). *A Standard Framework for Web Personalization*. Retrieved on 03.11.2008, from 1st International Workshop on Innovations In Web Infrastructure (IWI 2005) : http://www.wmin.ac.uk/~courtes/iwi2005/thomson_personalization_revised.pdf

[34] Wang, C.-Y., Wu, Y.-H., & Chou, S.-c. T. (2008). Toward a Ubiquitous Personalized Daily-Life Activity Recommendation Service with Contextual Information: A Services Science Perspective. Hawaii.

[35] Eirinaki, M., & Vazirgiannis, M. (2003). Web Mining for Web Personalization. *3*, 1–27. (ACM, Ed.)

[36] Mobasher, B., Cooley, R., & Srivastava, J. (2000). *Automatic Personalization Based on Web Usage Mining*. Retrieved on 09.04.2009, from DePaul University - College of Computing and Digital Media: http://maya.cs.depaul.edu/~mobasher/personalization/