SIM-Filter: User Profile based Smart Information Filtering and Personalization in Smartcard

Abdelkrim Lahlou[1], Pascal Urien[1,2]

[1] Schlumberger Systèmes, Smartcard Research Center, 36-38, Rue de la princesse, BP 45,
78431 Louveciennes Cedex, France
Abdelkrim.Lahlou@louveciennes.sema.slb.com
[2] Ecole Nationale Supérieure des Télécommunications - ENST Paris, 46, Rue Barrault
75634 Paris Cedex 13, France
Pascal.Urien@enst.fr

Abstract. The emergence of new information and communication system technologies allow the appearance of new techniques of data processing. Thus, it is increasingly difficult to user to find pertinent information in reasonable duration, unless knowing what exactly is looking-for, where and how getting it. We propose an Internet based filtering system that integrates user profile (i.e. identity and preferences data). User profile is embarked in a smartcard in order to guarantee its privacy and confidentiality. Our system objectives are twofold. First, we propose an evolved and extensible model to represent the user preferences with guaranty of privacy and security. The more the user surfs on the web, the more his profile is enriched implicitly. Second, our system must provide the pertinent information to the user according to his profile at any time, anywhere, and in any form. The design and implementation of our system is also presented.

1 Introduction

The advancement of networking and Internet technology has recently driven the rapid development and spread of World Wide Web. The emergence of new information and communication system technologies allow the appearance of new techniques of data processing. The services providers deploy new accessible personalized services through the Web that becomes vector of fast and effective communication. Not only, many information sources become available, but the exponentially information expansion also raises another subtle problem. Thus, finding relevant information effectively on the Internet is a challenging task. The informational potential made available to every one is becoming so considerable that it is increasingly difficult to user to find pertinent information in reasonable duration, unless knowing what exactly is looking-for, where and how getting it. Therefore, although searching tools ease the finding of information, the main problem to explore Web sites remains. What the user really need is a software agent that provides him the available and pertinent information at any time, anywhere, and in any form. Realizing such a promise depends on innovations in areas that impact the creation of information services and their communication infrastructures. Therefore, in this paper, we present an Internet based filtering system that uses user's profile (user's identity and preferences data). This system provides to users pertinent information and guarantees user privacy and confidentiality. User profile is embarked in a smartcard, which provides mobility and high security.

User profile can be used in many contexts among which information personalization and mobile computing. This information contained in the user profile is very confidential. For this reason, securing user profile is crucial. These points are presented in what follows.

(1) User profile and information personalization: The complexity of information research criteria constitutes the principal difficulty for the users when they formulate their requests, in particular when they specify the same criteria in a recurring way in various requests. The modeling of these criteria in the form of profiles allows mitigating these problems. It is possible for the user to partially specify his requests, which will be automatically supplemented and enriched by his profile before their execution. The user's profile can be defined like a modeling of his centers of interests (spheres of activity, level of data quality, level of security, etc). It can be described by a schema (metadata) and whole of request-types. The profile modeling provides to the user a simple way to express his requests. These requests can be rewritten and enriched by user profile before their execution. The rewriting mechanism is close to the rewriting of request through the databases views that is treated in an abundant way in databases literature [1], but with a higher degree of complexity. The information filtering and personalization in the context of information retrieval of text, audio or video type, such in the search engines or digital libraries, is out of the scope of this paper. Nevertheless, we will be inspired of the later works, in particular how to model user's profile.

(2) User profile and mobile computing: In the context of mobile computing, the user mobility entails various problems bound to the limits of the current storage and communication infrastructures. On one hand, the user is equipped with a mobile terminal presenting some limits such as: processing capability, storage capacities, energy autonomy, weight and dimensions, working interface ergonomics, etc. On the other hand, he is constrained by limited characteristics of the communication support such as: reduced bandwidth, raised disconnection ratio, various radio signal disturbances, variations in performances and reliability, data security vulnerability, etc. The mobile user wishes to find a navigation interface similar to that he has at home, in work or elsewhere during his mobility [2]. He hopes that his work environment will be reconfigured automatically at each connection. The user wishes also to maintain and synchronize his mobile terminal information with fixed data sources to which he accesses. He would get accurate information at time of his research and taking into account his preferences and his actual or future situation. These constraints make even more complex the task of managing the nomad user mobility. In this context, the personalization of the user interface according to his profile and constraints of the mobile environment (mobile terminal and radio interface) will be very helpful. It is also essential to guarantee the security and the confidentiality of his data and requests with his displacements. In our context, and since the user profile is embarked in the smartcard, the user can find easily his execution environment.

(3) Securing user profile: The representation of the entire user personal information, such as his services and presentation preferences, makes them very vulnerable. This personal data requires a high security level so that the user relies on his service/information providers. The confidentiality and the integrity of this information must be guaranteed. Therefore, a mutual confidence between the user and his providers must be established. The loading of these data in smartcard constitutes a high security level while making them mobile. As we present after, a judicious choice on representation of the profiles in smartcard is necessary, considering its limited characteristics about computing power and storage capacity.

Our application scenario is shown in Fig. 1. It represents our Internet based filtering system "SIM-Filter". Our system objectives are twofold. First, we propose an evolved and extensible model to represent and exploit user profile with guaranty of privacy and security. Second, our system must provide the available and pertinent information to the user according to his preferences at any time, anywhere, and in any form. The user profile contributes to offer a personalized information search and adaptable diffusion services to the users. It also helps to propose to users the most appropriate information, in suitable form and minimal effort. The user introduces his personal profile (identity and preferences data) in an explicit manner via a Java Interface. He can access to his profile anytime to reach and update its content. The profile is enriched also in an implicit manner according to the user's requests and selected responses. The goal of the implicit maintenance of the profile is to make it evolutionary and extensible. Our system integrates the user's profile in his requests before they will be processed. The more the user surfs on the web, the more his profile represents him and then the more his results are relevant. We design and implement our system on adapted HotJava[1] browser and we embark the user profile on Java smartcard.

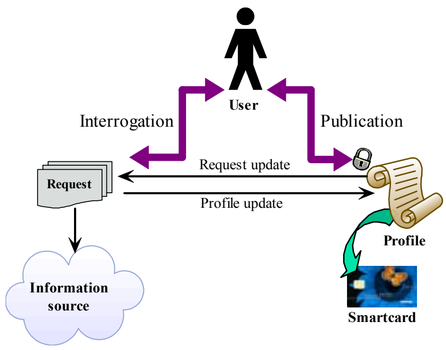

Fig. 1. SIM-Filter components

The rest of this paper is organized as follows: Section 2 reviews some related works on information filtering and user profile modeling. In Section 3, we present how to embark user profile in smartcard. We describe smartcard technology issues,

[1] http://java.sun.com/products/hotjava/

and user profile building and implementation in smartcard. We describe our Internet based filtering system architecture and present an execution scenario in Section 4. We conclude and present our future work in Section 6.

2 Related Works

A filtering is crucial task in information delivering to users. User's needs are modeled by his profile. Many works have been done on information filtering and many techniques have been proposed. This section presents some related works on information filtering and user profile modeling.

2.1 Information Filtering

Belkin and Croft mention in [3] that information filtering describes a variety of processes involving the delivery of information to people who need it. Although, this term is appearing quite often in popular and technical articles describing applications such as electronic mail, multimedia distributed systems, news services systems and digital libraries. Filtering applications typically involve streams of incoming data, either being broadcast by remote sources such as news servers, or sent directly by other sources such as E-mail. Filtering concerns also the process of accessing and retrieving information from remote databases, in which case the incoming data is the result of the database requests.

Information filtering systems are based on description of individual or group information preferences, often called profiles. Such profiles typically represent long-term interests of the user or community of users. The filtering process is often meant to imply the **removal** of data from an incoming stream, rather than **finding** data in that stream. In the first case, the users see what is left after the data is removed; in the latter case, they see what is extracted after the data is found. A common example of the first approach is an electronic mail filter designed to remove "junk" mails. Whereas, in the second approach, an example is to find only mail coming from a particular "diffusion list". This means that profiles may not only express what users want, but also what they do not want.

There are different filtering techniques referred in several research papers such as Boolean filtering [4], vector space techniques [5],[6],[7],[8],[4], probabilistic filtering [7],[9], linguistic and learning filtering techniques (neural networks, Case-Based Reasoning, genetic algorithms, etc.) [1],[11],[12],[13],[14], and [15].

2.2 User Profile Modeling

The Platform for Privacy Preferences Project (P3P), developed by the World Wide Web Consortium (W3C), determines a global architecture to preserve confidentiality on the Web [16]. This architecture offers secured profile exchange between a customer and a server. The goal of P3P is twofold. First, it allows Web sites to present their data-collection practices in a standardized, machine-readable, easy-to-locate manner. Second, it enables Web users to understand what data will be collected by the sites they visit and how data will be exchanged and used. The Web site accesses using a script, in reading state, to user's private preferences with respect of specific confidentiality constraints. On the other hand, all access attempting, in writing state, to profile is rejected. The P3P profile advantage is to (i) represent the user's preferences data, (ii) share these personal data with Web sites and (iii) guarantee their privacy and confidentiality. However, the P3P profile contains only static information and limited to the user's identification without taking in consideration his preferences. In P3P, the user can introduce or modify directly his profile data via an interface. The profile modeling is simple because the user's personal data does not frequently evolve. It does not take in account the dynamic aspect of the profile (i.e. preferences).

A concrete example of using user profile is presented in [2]. The authors present and implement a general user profile model in the context of digital libraries. The system is a personalized gathering and delivering for heterogeneous information sources based on user profiling model.

It is clear that except the P3P project, the majority of the systems and existing tools of filtering disregard the security aspect of the user profile. However, P3P limits itself to the user's identification without taking in consideration his preferences. It does not also permit to take in account the dynamic aspect of the profile; it is limited to a static profile.

We extend the P3P model to represent user profile by including its dynamic aspect (i.e. user's preferences). The user introduces his preferences in an explicit manner. The profile is enriched dynamically in an implicit manner. Fig. 2. shows user profile components. Our Internet based filtering system exploits user profile to remove irrelevant information and to find pertinent one.

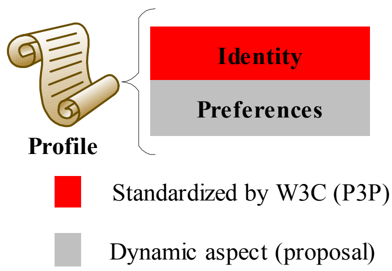

Fig. 2. User profile components

3 User Profile on Smartcard

Smartcards are the most secure portable computing device today. They have been used successfully in applications involving money, proprietary and personal data (such as banking, healthcare, insurance, etc.). The integration of the smartcard in information filtering process guarantees higher security but it presents two main constraints: (1) performances in terms of computing capacity and (2) storage capabilities. In this section, we present first the smartcard technology and then the implementation of user profile in Java smartcard.

3.1 Smartcard Technology

A smartcard as represented in Fig. 3. contains most of classic computer components, i.e. a CPU, a ROM (storing operating system code), a RAM (dedicated to stack operations), a persistent EEPROM memory (storing java byte code or user data), a communication bus, and a way to communicate with the outside world via an Input/Output connector. Silicon technology has considerably reduced the transistor size for smartcards to reach 0.18-micron process thus enabling important integration progress. The CPU of current smartcards is a 32-bit microprocessor with a processing power of 33 MIPS (Million Instructions Per Second) at a frequency of 33MHz.

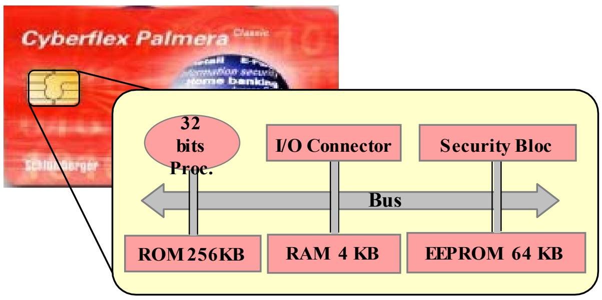

Fig. 3. Smartcard global architecture

The smartcard components need 50µs to process a cryptographic DES algorithm and only 200ms are necessary to compute a 1,024-bit RSA signature. As mentioned in [17] crypto-coprocessors could be replaced by dedicated cryptographic instruction in CPU cores. Memory capacities range from 128 to 256Ko of ROM, from 64 to 128Ko of EEPROM and from 4 to 8Ko of RAM. Writing data in EEPROM is relatively slow. It takes 1ms to write a 32 or a 64-bits word and it can be done only one million of times. The performance of smartcard's components will be a key issue in a closed future, as it could allow smartcards to be considered as the most secure communicating object on the Internet. Smartcard's capabilities are drastically limited due to its size. Fortunately, this will evolve rapidly. Smartcards with memories of few megabytes have been announced for 2003. Access to those memories is also a problem that will no longer exist with new memories such as FeRAM (10^9 writing operations allowed, memory capacity around one megabyte and writing delay less than 200ns). Processors frequency will probably progress according to the Moore law. One of the major chip manufacturers foresaw that the CPU would only occupy 10% of the area on the next generation of chip leaving more space for bigger Non Volatile Memory. Finally, one could expect secure dies with computing performances around 200 MIPS and memory capacities about 1 Mb, at the horizon of 2005's.

3.2 User Profile Building

User profile is built using information personalization techniques. We distingue two kinds of personalization: the explicit and implicit personalization. Explicit personalization consists in interrogating user directly on his needs, whereas, in implicit personalization, the profile is built by examining (in real time or not) user operations on the Web. The **explicit/implicit personalization** is also called personalization or **filtering by rules**. Our system is based on information personalization. In the both case of personalization, there is an automatically collection of information regarding user behavior. This information must be analyzed and classified according to the type of data viewed (consulted documents, clicks, bought products either wanted, Web search, etc.), then stored in a database. This perpetual enrichment permits to have a unified view of user behavior at anytime. Details about these two kinds of personalization are described in what follows:

(1) Explicit Filtering: In the explicit personalization (namely explicit filtering), the profile content is built according to the information given voluntarily by the user at the time of his subscription. The user describes himself, in explicit manner, his personal information and his centers of interests, via a form for example. It is called declarative or static modeling of user preferences. The explicit filtering presents the drawback to require the user's active involvement to describe his preferences. Besides, the user hesitates to introduce truthful personal information.

(2) Implicit Filtering: In the implicit personalization (namely implicit filtering), the profile content is built according to user's habits of navigation and while analyzing his clicks. The purchase forms, electronic subscriptions, demands of documentation are, as many others opportunities, a source to collect some information on user. The implicit personalization provides dynamic (evolutionary and extensible) modeling of user preferences. The implicit personalization presents the advantage that it does not require an active involvement of the user.

The combination of the explicit and implicit personalization techniques proves to be more efficient while integrating training techniques or the intelligent agents. These techniques put up to date or revalue the user profile according to his actions.

3.3 User Profile Implementation

User profile is well presented in a database. However, smartcards have severe hardware limitations (very slow write, very little RAM, constrained stable memory, no autonomy, etc.), which make traditional database technology irrelevant. The major problem is scaling down database techniques so they perform well under these limitations. In paper [18], authors give a depth analysis of this problem and propose a **PicoDBMS** solution based on highly compact data structures, query execution without RAM and specific techniques for atomicity and durability. We use this solution to implement user profile in the smartcard.

The user enters a new profile or edit/delete an old one via a java Interface (see Fig. 4.). This profile is stored on the pico-Database. The application uses a JDBCDriver and SQL statement similar to one used in a classical Database. It is completely transparent to the application to access the pico-Database with the JDBCDriver. We recall, that user profile is composed of user identity and preferences. Preferences are first stored in explicit manner, and then enriched implicitly. Our system observes the user's actions and deducts some of implicit data that represent his opinions on the proposed information.

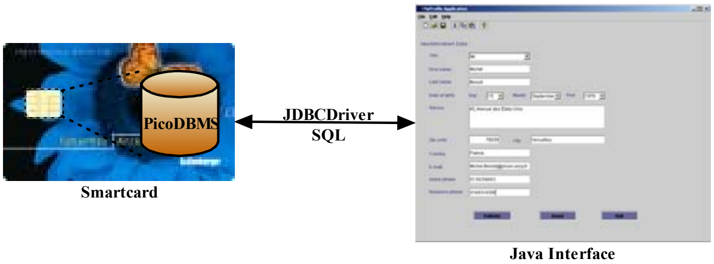

Fig. 5. User profile implementation

For example, suppose that user likes classical music. He setup his preference according to this choice. For this, our application maintains a list of MIME file type that specifies the type of each file in the Internet. When user clicks on music file on the web, we can detect the kind of this music (technique known as "click-stream tracking"). By analyzing the behaviors of the user, based on the number of time that a given link has been selected of a list of results, the number of time that a document has been consulted, the time passed on a document or a site, etc., we detect that the user often listen to rock music. Therefore, our system adapts user preference and update/add rock music to user preferences.

4 Internet Based Filtering System

There are two techniques of information filtering. Fig. 6. explains these techniques. In the first case, the user formulates his request (Request Q1), which is combined with his profile before queering the web. In the second case, the result of the request (Request Q2) is combined with the profile to get only pertinent response. The advantage of the first technique is the optimization of the resources used such as bandwidth consumption and execution time. It presents some drawback related to the security and confidentiality of the profile. It is clear, that in this case everyone can discover user profile, which may contain personal and private information about the user. It is necessary then to secure the storage of the profile (this is already provided by the smartcard) but also its transfers (network communication). The second technique deals with the security problem, since the profile is combined locally with the request only after getting responses. The drawback of this technique is the resources consumption.

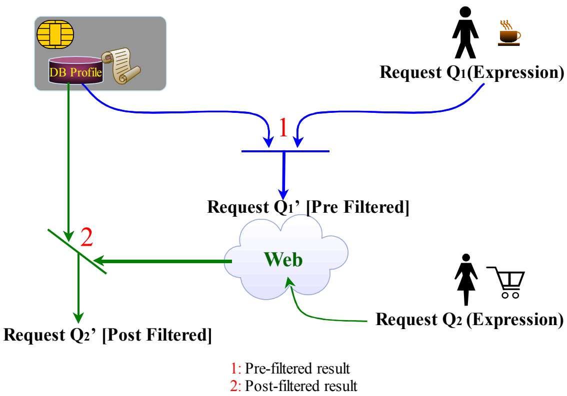

Fig. 6. Two Scenarios of Internet Filtering

4.1 System Architecture

Our implementation consists of Java module, which extends the functionalities of HotJava browser. HotJava™ browser provides a highly customizable modular solution for creating and deploying Web-enabled applications across a wide array of environments and devices [19]. Our module communicates with the smartcard using JDBCDriver. The smartcard plays the role of proxy since it can allow or not the browser to make connection in the Internet [20]. The architecture of our Internet based filtering system "SIM-Filter" is shown in Fig. 7. In reality, there are two execution modes. In the first mode, the smartcard authenticates the user's URL. For example it checks if the tapped URL is not forbidden. If the URL is acceptable then it allows the browser to **connect directly** to the Internet, else it sends a message error indicating that the URL is not authorized. In the second mode, the smartcard is viewed as proxy, in which the browser **connects indirectly** to the Internet (via the smartcard proxy). The last mode is not implemented in our system because of the limited capabilities of today's smartcard. All traffic going to the browser must be filtered in the smartcard, and this requires a high degree of computation.

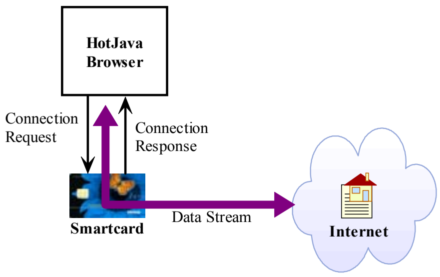

Fig. 7. Internet based filtering system architecture

4.2 Execution Scenario

Let us take the example of user who wants to find some flight (see Fig. 8.). The user specifies in his profile that he is only interested by flight with price lower than $300. When the user makes his request, the profile is combined with his query as shown in Fig. 6. This example presents only pre-filtering mode, which is actually implemented in our system. Post-filtered mode is more complex and requires high computation capabilities from the smartcard.

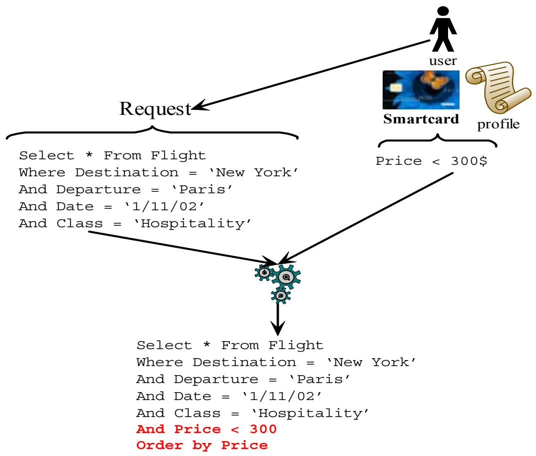

Fig. 8. Example of flight request

Conclusion

Finding effectively relevant information on the Internet is becoming a challenging task. In this paper, we presented a secure and extensible model to represent user profile. We extended the P3P model of identity representation by integrating user preferences. These preferences are introduced in an explicit manner and increase dynamically implicitly. Furthermore, embarking user profile in smartcard makes it more robust and high secure. We developed an Internet based filtering system SIM-Filter, which benefits from user profile to filter information. User profile can be combined before the request and is called pre-filtering model, or after the request and is called post-filtering model. The post-filtering model requires high computation capabilities, which are left on today's smartcards.

Actually, we are planning to provide an extended secured personalized filtering service based on our architecture on the context of Pay for view TV. User defines allowed TV channels and emissions. According to this information, our system can authorize movie view or not.

References

1. Halevy A.Y. : Answering Queries Using Views : A Survey. In VLDB journal, (2001) (http://www.cs.washington.edu/homes/alon/site/files/view-survey.ps)
2. Cherniack M., Franklin M.J., Zdonik S. : Expressing User Profiles for Data Recharging. In IEEE Personal Communications, (2001)
3. Belkin N.J., Croft W.B. : Information filtering and information retrieval : two sides of the same coin. In Communications of the ACM. Vol. 35(12), (1992) 29-38
4. Yan T.W., Garcia-Molina H. : The SIFT Information Dissemination System. In ACM Transactions on Database Systems, (2000)
5. Deerwester S., Dumais S.T., Furnas G., Laundauer T., Harshman R. : Indexing by Latent Semantic Analysis. In Journal of the American Society for Information Science, (41), (1990) 391-407
6. Foltz P.W., Dumais S.T. : Personalized Information Delivery: An Analysis of Information Filtering Methods. In Comminucations of the ACM, 35(12), (1992) 51-60 (http://www-psych.nmsu.edu/~pfoltz/cacm/cacm.html)
7. Larson R.R. : Evaluation of retrieval techniques in an experimental online catalogs. Journal of the American Society for Information Science 43(1), (1992) 34-53
8. Salton G., McGill M.J. : Introduction to modern Information Retrieval. McGraw-Hill Edition, New York, (1983)
9. Van Rijsbergen C.J. : Information Retrieval. Butterworths 2nd Edition, London, (1979)
10. Aamodt A., Plaza E. : Case-Based Reasoning: Foundational Issues, Methodological Variations and System Approaches. AI Communications, Vol.7, (1994) 39-59
11. Boughanem M. : Les systèmes de recherche d'informations : d'un modèle classique à un modèle connexionniste. Phd Thesis, Computer Science. Paul Sabatier University, Toulouse, France, (1992)
12. Goldberg D. : Genetic Algorithms in Search, Optimisation, and Machine Learning. Addison-Wesley,Wokingham, England, (1989)
13. Haykin S. : Neural Networks - A Comprehensive Foundation. Prentice Hall International, (1994)
14. Kobsa A. : User Modeling in Dialog Systems: Potentials and Hazards. AI & Society: The Journal of Human and Machine Intelligence, 4(3), (1990) 214-240
15. Shardanand U., Maes P. : Social Information Filtring : Algorithms for Automating Word of Mouth. Proceedings of the CHI'95 Conference, (1995)
16. P3P : Platform for Privacy Preferences Project home page. (http://www.w3.org/P3P)
17. Urien P., Tizraoui A., Loutrel M. : SIM-IP card benefits for a profitable business model in emerging VoIP technologies. In Proceeding eSmart'02, (2002)
18. Pucheral P., Bouganim L., Valduriez P., Bobineau C. : PicoDBMS: Scaling down Database Techniques for the Smartcard. In VLDB Journal, 10(2-3), (2001)
19. Sun Microsystems: HotJava Browser Home page. (http://java.sun.com/products/hotjava)
20. Urien, P. : Internet Card, a smart card as a true Internet node. Computer Communication, Vol. 23 Issue 17, (2000)