Information Personalization in an Internet Based Travel Support System

Violetta Galant
Department of Computer Science and Management
Wrocław University of Economics
Ul. Komandorska 118/120
53-345 Wrocław, Poland
violetta.galant@ae.wroc.pl

Marcin Paprzycki
Department of Computer Science
Oklahoma State University
Tulsa, OK 74106, USA
paprzyc@okstate.edu

Abstract

Human travel agents are one of the sources of conceptual metaphor for the development of software agents. It is thus natural that software agent based travel support systems have been introduced [Ndumu 1998, Paprzycki 2001a and 2001b]. These papers were concerned mostly with the technical issues related to the agent technology and system development. The aim of this paper is to address the questions related to the personalization of content delivery in the context of Internet-based travel support system.

1. Introduction

Information personalization is one of more pressing needs in the development of a human-friendly interface to the Internet. As the Web became the largest repository of information, the very vastness of the available data is also one of the reasons for its limited usability. Anyone who has searched for information using either of the search engines and received response that there are 234,567 pages containing "relevant" content understands the nature of the problem. One possible way to respond to this situation is the development of *personalized information services* (*PIS*). However, this task is facing a number a potential problems, which are very difficult to solve in the general case (see Section 2). Our goal is a limited one, to consider the *PIS* delivering travel-related information. Work presented here is an extension to the proposal presented in [Paprzycki 2001a and 2001b]. There, the general agent-based structure of the travel support system was presented [Paprzycki 2001a] and followed by a discussion focused on the server-side of the system [Paprzycki 2001b]. The proposed system is to support needs of travelers by fusing a Geographical Information System and a Business Information System. As always, the core of the travel support should be a map, and thus the geospatial information (and the *GIS*). However, today, having only a map is clearly not sufficient. Travelers prefer to have the geographical information combined with the information about *services* available in a given area (here the definition of services is a very broad and contains, among others, restaurants, movie theaters, museums, national parks etc.). Therefore, travel agencies, especially these residing on the Web have to be able to satisfy these expectations. In the optimal case, the information about services should match individual users' interests.

At the same time, the information personalization is also very important for the business aspect of Internet-based endeavors. It has been established (through the "dot.com industry" collapse in 2001) that the delivery of banner ad's targeting "everyone" does not represent a very viable business model (there are very few businesses that can utilize it successfully, and usually only as a supplementary source of income). Similarly, users turned out to be relatively resistant to the "push technology" (which seems to irritate, rather than attract, the potential customer). As a response to the growing amount of advertising, banner-ad/pop-ad filters have been developed [Dreyfuss 2002]. This seems to indicate, that a new business model for Web-vertising is needed. It is our belief that the 1-to-1 marketing strategies that attempt to deliver targeted advertising content, based on the knowledge about the user, may be one of possible solutions to this problem.

Our paper is devoted to discussing various aspects of information personalization in the context of an Internet-based travel support system (an "ultimate" Internet travel agency). Note that our paper is written from the "business perspective" and the aim of proposals discussed here is to increase the profitability of the endeavor. Therefore, while in the next section we briefly discuss various aspects of Internet content personalization, the remaining parts of the paper are devoted to the personalization approached from the business side of the e-commerce relation.

Witold Abramowicz (ed.), Business Information Systems, Proceedings of BIS 2002, Poznań, Poland

We proceed as follows. In Section 2 we introduce basic issues involved in information personalization and proceed to consider some of the important obstacles to the development of personalized information systems. Section 3 sketches the current state of the Internet-based travel-oriented services. We also briefly argue why a multisource data-fusion is necessary for a successful functioning of the system. In Section 4 we sketch the general architecture of the system and point where the personalization takes place. Section 5 contains the discussion of possible personalization techniques and how they are to address problems noted in Section 2. Future research directions and concluding remarks can be found in Section 6.

2. Delivery of personalized information

2.1 General considerations

When considering personalized content delivery, one can observe that this situation involves two sides: the information consumer and the information provider. The information consumer would like to obtain the desired information and **only** the desired information. Any additional content (e.g. advertising) is considered a price that one has to pay for the information received. On the other side there are two types of content suppliers. First, governmental and non-profit organizations that **provide** information to the users (without an explicit goal of selling anything). Second, businesses that attempt to **sell** goods and/or services to users and use the personalized content to generate sales.

Let us now look at the Web-based systems and consider content personalization. When visiting any portal one can characterize its overall quality by considering three factors. First, the way that the site looks/"feels." Second, the quality (and quantity) of the informational content. Third, the quality and quantity of the advertising content. These three characteristics translate into three categories of information personalization. (a) Personalization of the "way" that the information is presented/delivered. It includes: arrangement of the browser screen, colors used in it, the way in which the new content notification is send to the user etc. (b) Personalization of the informational content. Here we are interested in the way in which the information is filtered to match the user interests. (c) Personalization of commercial content. Here we consider all techniques resulting in 1-to-1 marketing.

Obviously, there is a natural match between the interests of the two sides of the content delivery and consumption relation and the three types of information personalization. The information consumer is interested in types (a) and (b) of information personalization and considers personalization of type (c) a price to be paid for the content she desires. The governmental and non-profit organizations may be ready to provide user with personalization of types (a) and (b) as a method of fulfilling their obligations to the citizens (governmental organizations), or a vehicle of spreading the message about their goals and aspirations (non-profit organizations). Let us now look in more detail at the role of all three types of content personalization in case of for-profit organizations (for more details see also [Chaudhury 2001, Lohse 1998, Mohr 2001, Strauss 2001]).

From the e-commerce perspective, one of the more important goals is to generate repeated site traffic and building a lasting relationship with the customer. Information personalization is one of the more powerful tools to achieve this goal. Each of the personalization types described above plays an important role in achieving this goal. Type (a) personalization helps tailor users' browsing experience by matching their specific characteristics and preferences. This makes the system attractive to the users (in particular to the sophisticated "power users" see also [Manber 2000]). Type (b) personalization provides users with the "value-added" information[1]. While it does not necessarily lead directly to sales, it should keep customers interested in returning to the system. Each such repeated visit may result in a transaction (even if one was not originally planned). Type (c) personalization is the core of the revenue stream for the Web-based systems.

Observe that the repeated traffic is not only directly beneficial, due to the fact that it may result in sales. Each visit allows the system to gather additional data about the customer. This helps to further develop the user profile. This, in turn, improves the quality of personalization (of all three types).

2.2 Potential problems

While the support for type (a) personalization is relatively well developed and with some effort could be further improved (for instance, a substantial part of Microsoft's .Net strategy clearly leads in this direction), the remaining two types of content personalization face a number of potential problems. We will present seven that seem to be of particular importance (this list is not considered exhaustive, but we believe that the most important issues have been included).

[1] We will omit here the discussion of the important issues related to the possible abuses of the system (e.g. businesses that adapt the non-advertising content to avoid directing attention of the reader to other companies, or to demean competitors). While very important as an issue in its own rights, it clearly falls beyond the scope of this paper.

(1) When the user visits the site for the first time, the system does not know him. The *PIS* works correctly and is effective, if it has some information about the user (obviously, the more information is available and the more effectively it is processed into the content delivery rules, the better the results). One of the ways to obtain the initial information would be to ask the newcomer to fill in a questionnaire. However, nobody likes to fill long, time-consuming questionnaires (and the Internet seems only to amplify our desire for instant gratification). How are we then to initialize the interaction with the new customer in such a way to be able to keep him? This problem is often addressed by utilizing rules based on some form of "weighted average preferences." For instance, many search engines use this approach and rank the returned pages based on the information about how many times other visitors accessed them in the past in return to a similar query. Similarly, "zero-knowledge systems" attempt at suggesting information to their user solely on what other users considered in a similar situation [Lieberman 2001]. Unfortunately, at a certain level, this approach is working in the opposite direction that the personalization should.

(2) In the case of personalization taken seriously, we are not interested in what "everyone" thinks the result of the query *X* should be, or what everyone would click on a given page. Rather, we are interested in providing the individual user with the information that she is interested in (regardless of how far her preferences may be removed from the popular views). This situation causes problems in at least two ways: (a) the process of adjusting the *PIS* to the non-standard user may take a long time (system has to learn about the preferences that are not represented among existing rules, and that are very different from these that already are represented there) thus discouraging the user that we are trying to sway to stay with our service, and (b) for each non-standard user a new and different set of rules needs to be derived and stored, thus consuming system resources. While problem (b) should be solvable by hardware and software upgrades (and thus is purely technical in nature), it is problem (a) that, due to its conceptual nature, is of much greater concern.

(3) Let us now assume that we were able to convince the new client to use the system. In this case we need to keep in mind that his preferences are likely to change over time and it is almost impossible to make any a'priori prediction about the direction of such a change (e.g. a person who was once interested in primarily in visiting dance clubs starts visiting art galleries). Such a change can be a slow process, and in this case the system should be able to adjust its behavior, or it can be an abrupt change (e.g. caused by client entering a relationship with an art critic). Note that this problem is typical for any case of an extrapolation from the present to the future. It is only when the behavior remains relatively constant, or when it changes at a moderate rate the prediction methods can be expected to work well.

(4) It is also possible that while the core set of preferences remains constant, users may enter the system in different moods (in love, depressed, high on caffeine etc.) or simply become bored and interested in exploring new areas (the "I am in a mood for something different tonight" syndrome). A *PIS* that is very well tuned to the user does not match with human curiosity and the desire to explore, as it provides information limited in scope to the fixed set of recognized interests (one can say that in some sense the system is "over-trained"). Observe also, that it is exactly the dreaded generalized browsing, where at any time the user can step into a new unexpected area of the Web that conceptually matches human curiosity better.

(5) Problem of sudden interest changes can occur for each individual user. However, there exist situations when events taking place "outside" of our system can significantly change the behavior of all, most, or a particular group of users. For instance, the Olympics may result in increased traffic to Salt Lake City (and increased interest in Utah and/or Mormons) among a large subset of potential travelers, whereas the expected eclipse of the sun over Northern Alaska is likely to generate interest in a highly selective group of individuals (note that being able to **correctly** target these individuals with the promotional offers is the ultimate goal of a system like the one discussed here). Finally, a sudden travel advisory issued by the Department of State asking US citizens to stay away from Nigeria should, among others, result in the system contacting all users who have already purchased vacations there. Therefore, external influences have a direct effect on the personalized content delivered to the users.

(6) We need also to pay attention to the serious privacy concerns that can be associated with any system that gathers data about the user behavior. It is quite possible that the client may not want that information about his behavior be collected and/or stored. For example, the *PIS* may collect information about amount and brands of alcohol that the client drunk in the hotel. While such an information may be useful to the system (it may

allow to introduce targeted marketing consisting of special offers for the alcoholic beverages matching clients "interests"), the fact that this information was collected may be perceived as an abuse of trust and result in clients abandoning the system. Because of the concern for the preservation of their privacy, some people possibly will also try to shield their identify and personal facts. In this situation the *PIS* may never reach the full potential that the personalization can provide.

(7) Purposeful abandonment of the system may be only one of the reasons for loosing customers. Users often forget theirs log-ins and/or passwords. While in most cases it is possible to recover such information, sometimes, for security and privacy reasons such a recovery may be impossible (e.g. the user does not have an e-mail account that was associated with his name anymore, and thus we cannot send him the password information). It may seem that the re-registration to the system is an easy and fast solution. But for the *PIS*, the re-registered customer has to be treated as a completely new one, with a blank record. The system must therefore collect information about this customer again (which clearly is not an activity that the returning user will enjoy). At the same time, in the customer database there will remain stale data, which will never be used again (and has to be discovered and dealt with).

Summarizing, from the business point of view, winning the new client is a complicated and expensive long-term process [Strauss 2001]. For this reason the initialization of a user profile represents a non-trivial task of in the development of the *PIS*. However, problems do not stop there. To keep the client, her needs have to be satisfied to some reasonable extent and to achieve this goal the system needs to be truly personalized, while still flexible and responsive to the privacy concerns. We will try to address these points in the context of the travel-oriented *PIS*. We will start our discussion with a summary of the state-of-the-art in travel information delivery and personalization. We will follow with a brief introduction to the agent based travel support system co-developed by one of the authors in [Paprzycki 2001a and 2001b]. We will use this information as a backdrop against which we will discuss the personalization of travel information delivery.

3. Internet travel services today

The need for information personalization is well recognized and most large portals support some form of customization of Web experience for individual users or groups of user (e.g. three versions of the CNN portal: European, Asian and US oriented www.cnn.com, or a number of interesting personalization tools offered by the Yahoo! Portal, www.yahoo.com [Manber 2000]) attempting to match content and the way it is displayed with users' preferences. It can be almost stated that (at least for the large portals) the "value" of the Web site content can be "measured" by the level of available personalization. At the same time, personalization seems to be somewhat neglected in Web-based travel information services, where it is typically limited to a few preferences related to airplane seat assignment, car rental preferences and hotel/motel preferences (see, for instance, www.travelersadvantage.com). As noted above, the scope of services that travelers may be interested in is much broader than this. Obviously, one can find information about many (if not all) additional services, but they will be stored in separate repositories (e.g. the information about the local ZOO is not likely to reside in the same repository as the program of the local Opera). Finally, even a minimal contact with the Web-based travel sites allows one to realize that it is not a "democratic" place. Large companies get a lot of exposure, while the smaller players may be extremely difficult to find. We challenge anyone to find information about Hotel Batory in Kraków, Poland (without previous knowledge of existence of such a hotel) and compare the effort with that of finding the local Holliday Inn. A slightly different example would be the conflict between Southwest Airlines and the Orbitz portal (www.orbitz.com). Here, Southwest Airlines do not want their fares to be posted on the integrated airline reservation site (as they prefer customers to purchase their tickets directly from their own travel site). The latter situation is somewhat a problem for any attempt at building an all-encompassing travel support system. However, since it belongs to the considerations related to the e-business model, it falls outside of the scope of this paper. Here, we assume that all businesses elect to participate in our system and are treated equally when the travel-related information from multiple sources is integrated. For the purpose of this paper we will also assume that a unified repository of travel related information as suggested in [Paprzycki 2001a and 2001b] exists and use this as the context in which we will discuss the delivery of personalized travel information.

4. General system architecture

The general architecture of our proposed solution is illustrated in Figure 1, and its basic elements are: *content providers, Internet enabled devices, data federations and data personalization services.* We will only very briefly sketch the overall model of the system

and concentrate our attention on the information personalization (further details can be found in [Paprzycki 2001b]). Let us note here, that the overall structure of the proposed system is quite similar to that proposed in [Paralic 2001] (with application to retail support and to e-Democracy) and at least some of the tools and methods used there can be easily modified to out project (thus making it more feasible). Additionally, the agent-based personal travel assistant (*PTA*) described in [Ndumu 1998], while limited to making travel reservations travel arrangements, fits naturally into our project as it provides parts of the necessary functionality.

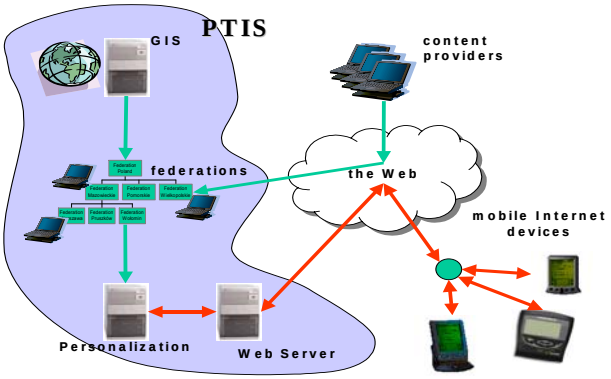

Figure 1. An overview of the Personalized Traveler Information System

Content Providers

Today, there exist a very large number of Web-based providers of information that can be useful for travelers. Here are some examples of such sites grouped into a few general categories (this list is only intended to illustrate the breadth of available information):

- weather information
  - http://www.pogoda.eu.org/
- information about dangers
  - hurricane tracking software provided e.g. by http://www.weather.com
- sites of interests – special tourist sites created locally and most large portals
  - www.yahoo.com
- accommodations
  - www.hotelspoland.com
- dining
  - http://www.restaurants.com
- timetables and fares
  - www.pkp.com.pl
- rentals
  - car – www.rentalcars.com
  - moped and bicycle – www.cwdesign.co.uk
- special forms of tourism e.g. agrotourism
  - http://www.agroturystyka.pl

Since the proposed system works as a bridge between content providers and users, it belongs to the category of infomediary systems [Grover 2001]. We will omit questions about "trustworthiness" of content providers, dealing with the conflicting information and the information validation as they are outside of the scope of this paper and assume that they have been successfully resolved.

Internet Enabled Devices

Access to the system will be facilitated via an Internet enabled device. In addition to the standard PC-based browsers, access can be obtained via Laptop computers, Wireless Application Protocol (WAP) browsers or palmtop units. Obviously, each device has its own advantages and disadvantages (in the context of travel related information delivery) and a partial discussion of these issues can be found in [Paprzycki 2001a and 2001b].

Federations and the GIS

In the proposed approach, the information provided by the content providers is *indexed* into federations according to the geographical location. Here, federations are the way of combining the geospatial data with the remaining travel-related information. The geospatial data, represented in one of the standard geo-representations, e.g. a quad-tree becomes the backbone of the travel-oriented hierarchical indexing structure. The scope of the federation may vary from the continent to the city museum. The schema depicted in Figure 2 illustrates a part the hypothetical section of the division of Poland.

Since a hierarchy of federations is used to fuse travel related information with the geographical information, separate data items provided by the content providers will be indexed in the lowest level federation that completely encloses their geographical extent. For instance, information that PLN is the Polish currency will be associated with the federation Poland, while a timetable of night buses in Warszawa will be associated with the federation Warszawa.

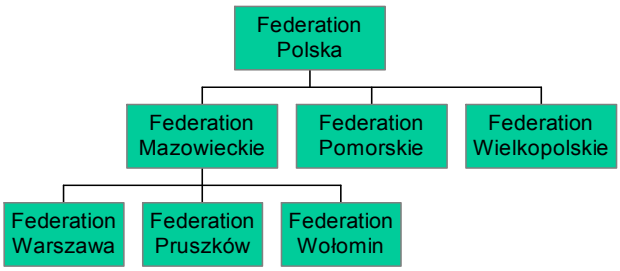

Figure 2. An example division of the information into federations

This approach is similar to the concept of *Location Organizer Folders*, introduced by the Open GIS Consortium [GIS 2001] and the methods developed in the context of this proposal can be naturally applied here. Similarly, the question about designing ontologies for Web-based applications has been recently addressed in the Special Section of the *CACM* [Gruninger 2001] and the techniques and tools described there will be very useful when implementing our system. Finally, while most of the current work is concerned with the delivery of text-based information (which is much easier to deal with and a lot of results already exist), the overall idea discussed here is easily and naturally extendable to the multimedia-type information. Therefore, as soon as the theory and practice behind management of multimedia content (e.g. search for photographs and/or movies) matures, the same approach to their usage in travel support can be applied. The discussion of advantages and disadvantages of systems build around the "indexing hierarchy" (storing only the information about the location of pertinent information) can been found in [Abramowicz 2002] and references presented there.

5. Personalization of travel information delivery

5.1 Structure of the personalization subsystem

We will start our discussion with the introduction of the general structure of the subsystem responsible for the information personalization (see Figure 1, for its location in the overall system structure)[2]. Let us assume that at a certain time the initial set of geospatial and travel data have been fused into federations and the system is "ready to be used." The proposed *content personalization system* consists of the following modules: *interface software, a number of domain specific expert systems, user profile database and user behavior database*[3] (see Figure 3). Let us now describe each of the proposed modules and then summarize their interactions.

Interface and management software

The *interface and management software* module combines all of the necessary the functions involved in human-computer interactions. At this stage of system development we envision that it will encompass four software subsystems, which closely interact with each other: (i) software that manages user accounts (account creation, deletion, password management, subscription management – in case of payment-based systems etc.), (ii) software that manages the personalizable user interface (variety of tools that are involved in the type (a) personalization), (iii) software responsible for users' interaction with the system (e.g. translation of user requests into system queries, personalized content delivery etc.), (iv) management software that interacts with the remaining components of the system (primarily management and scheduling of various actions performed by the *domain specific expert systems*).

User profile database

This database contains the static user profiles. Each user of our system has its own profile that is utilized by the *travel expert system* and the *marketing expert system* to deliver the personalized information and marketing content. The initial profile will be created using one of the methods described in Section 5.2. In addition, the profile will be occasionally modified using the knowledge discovered in the data accumulated in the user behavior database.

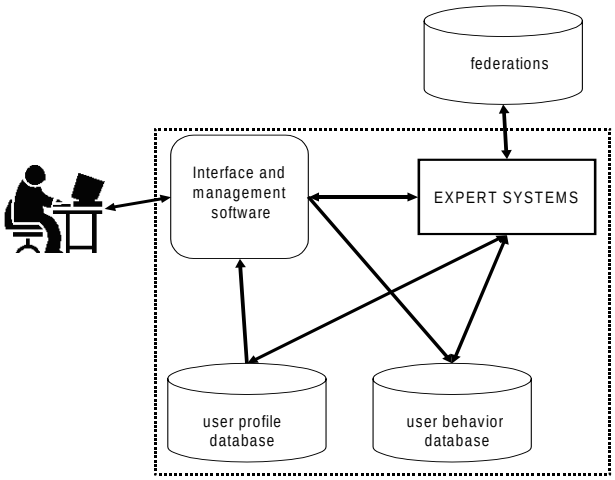

Figure 3. The general model of the *content personalization system* (inside of the dotted square)

User behavior database

In this database the information about the user activities is stored. There are two types of information that can be collected. First, only the information related to the interaction with our system is collected (and this data includes both the information about the completed transactions, as well as supplementary information, e.g. banner ad's clicked on, the non-commercial information requested, etc.). Second, information about the user behavior on the Internet (not necessarily related to travel) is also collected. While in the first case, the

[2] While the personalization is a *subsystem* within the overall travel support system, for simplicity we will refer to it as a *system* through the remaining parts of the paper.

[3] While these two databases can be implemented as separate tables in a single database, in our discussion, for the clarity of the picture, we will treat them here as functionally separate units.

server log files will be used, in the second case the cookie-like technology can be utilized. Data stored in the user behavior database allows us to study users behavior over time and can be used to supply the missing data into the user profile, modify the user profile, modify the user profile clusters, and to extract new rules that should be included in the specialized expert systems described below.

Domain specific expert systems

In the proposed system, ability to deliver the required functionality entails existence of a number of (domain specific) expert systems. While it is more typical that knowledge is embodied in a single expert system, we believe that such a situation is rather unnatural for the development of large complex systems. Following the software engineering principles introduced in [Jennings 2001], we plan to implement our system using agent technology (see also [Paprzycki 2001a and 200b]). In this approach, it is quite natural to embody multiple, domain specific expert systems into separate static agents, interacting with each other to achieve their goals. One of the primary advantages of this approach is the expected ease of maintenance of each expert system, which exists as a separate unit. The price to be paid is the need to design interfaces and interaction rules between these agent-embodied expert systems. Here we present the list of expert systems need for which we have established so far (it is quite possible that need for further expert-agents will arise during subsequent stages of system development and implementation):

(a) *User profile initialization expert system* will be used to create the initial user profile to be stored in the *user profile database* (see Section 5.2 for more details). Some of the techniques that are going to be applied during the account initialization can be also applied later, when the user profile will be refined using knowledge extracted from the *user behavior database*. We may therefore try to add the user profile modification functionality here, creating a more general *user profile expert system* (see also Section 5.3). The decision if the user profile initialization and user profile management functions should be combined, or kept separately, will be made in the next stage of system analysis and development.

(b) *Travel expert system* contains travel-related rules e.g., business travel often involves more expensive car rentals, first class airline tickets and more expensive hotels and restaurants, while the family travel involves renting a van, economy tickets with seat assignments that group family members together, cheaper motels and visits to local attractions. It is responsible for selecting the information to be delivered to the customer. This system will likely be initialized through interactions with human travel experts and will be further refined during the lifetime of the system, based on the collected data about the customer behavior (obtained through the data mining/knowledge extraction from the *user behavior database*). Using terminology introduced in Section 2.1, this expert system is responsible for type (b) information personalization.

(c) *Marketing expert system* is responsible for type (c) content personalization. It will contain all the rules related to the delivery of personalized marketing information, e.g. direct service advertising, general banner advertising, information about promotions and sales (see also [Brady 2002]). Similarly to the *travel expert system*, it will be created on the basis of interacting with human marketing experts, in its work will utilize the *user profile database*, and will be refined on the basis of the user behavior observed during the lifetime of the system (another use of data mining/knowledge extraction techniques applied to the *user behavior database*).

(d) Finally, our system will contain one or more of *meta-expert system(s)* that will be responsible for mining the data available in the *user behavior database* for additional rules that should be added to the expert systems described above. In addition, expert systems dealing with customer clustering etc. also belong to this general category[4]. This functionality is extremely important for the operation of the system. Since the initial set of rules will be extracted from human experts, they do represent only what these experts believe the best practice should be. We believe that for the system to successfully interact with the real customer population, it has to learn and refine its behavior to match that of the users. For instance, based on the user behavior, a new product association rule (the fact that honeymoon suite is associated with champagne purchases) may be discovered and added to the *marketing expert system* to target advertise with champagne specials these who make

[4] Obviously, we are using here a rather loose definition of an expert system and some of the tools mentioned here do not adhere to its strict definition. However, since this paper is concerned with information personalization, we have decided to forgo the precision for the sake of brevity hoping that this will not result in confusion.

reservations in honeymoon suites. Human experts may omit such an obscure rule, thus causing the system to miss a perfect marketing opportunity.

Let us now briefly summarize the expected interactions between the above-described modules of the personalization subsystem.

User interacts directly with the *interface software*. New users will be assisted by the *user profile expert system* that will generate the initial profile to be stored in the *user profile database*. Returning users will cooperate with the system to obtain the requested help with their travel arrangements. In this process they will use the *travel expert system* that will make their interaction with the system faster (due to the fact that the information stored in the *user profile database* contains user preferences, which can be effectively utilized in reducing the search space) and deliver better quality information (again, the user profile helps the system to find what the user is likely to be looking for). The *marketing expert system* will utilize the user profile and the user query to deliver personalized marketing information that matches both (e.g. family going to San Diego will be informed about the Sea World specials). After the interaction with the system is concluded, the complete report describing it will be stored in the *user behavior database*. This information will be then used by the *meta-level expert systems* to refine the user profile and search for new rules to be included in the appropriate expert systems. The data mining/analysis activities will be scheduled by the management software (attempting, for instance, to balance the workload of the system and thus scheduling the data analysis activities during off-peak hours).

5.2 User profile initialization

To discuss the delivery of personalized information we will proceed in the context of the seven problems introduced in Section 2.2. Let us start from the moment when a new user is signing on to the system. She is provided with a new account. There are four possibilities to initialize her user profile (see also [Brenner 1998]):

1. Explicit initialization

   Upon logging-in for the first time the new user is asked to fill a questionnaire(s) that describe her travel preferences. They contain, for example, questions about her favorite hotel chain (and room type, and bed size, and acceptable price range etc.) or the preferred airline (and the seat preferences, and the acceptable price range etc.). The system can then directly record the obtained data in the user profile database. Observe, that in this case we expect the user to have well-developed travel preferences (which may represent a problem in itself). The main disadvantage of this approach lies in the fact that the process of filling a detailed questionnaire(s) is time-consuming and rather boring and may discourage potential users.

2. Initialization using user characteristics

   The new user is asked to fill a special questionnaire with the information related to the selected characteristics (typical questions are related to: address, age, gender, occupation etc. [de Kare-Silver 2001], and answering them can be made a part of the account opening process itself). In this case, the *user profile initialization expert system* contains information about user classes (created using the data clustering software) and the received answers are used to match the new user to one of them. In the case of the travel support system, in addition to the above listed categories, very valuable information concerns the family status, income level, professional interests, hobbies, etc. Obviously, it is necessary to make the questionnaire rather concise. Three techniques should be applied here: (a) only a minimal number of questions is asked, (b) in case of longer questionnaires, only few questions are required to be answered (the additional questions are used if answered, however a lack of an answer is also acceptable and the user will not be required to specify more than the bare minimum) and (c) the newcomer will be offered to fill-in additional questionnaires, however their completion will not be required (we may apply the latter approach also to the returning users, by occasionally asking them to fill in additional questionnaires and offering them small rewards for providing us with the information; this information can then be used to enhance the quality of the user profile). The important advantage of this method is its simplicity. However, its precision depends on how accurate we were in creating groups of distinct user profiles, how many such profiles were we able to distinguish and how well the received answers match either one of them (the question of dealing with an eccentric customer comes here in full force). Here, there exist a number of trade-offs that have to be carefully considered to reach an optimal solution (for more details see [Brady 2002]). Let us note that the information stored in the *user profile database* is very valuable for future mining and may be used (separately, or together with the user behavior data) in a loop-back to modify the initial set of customer classes (see also [Rud 2001]).

3. Initialization using examples

   One of the known problems with creation of expert systems is the formalization and verbalization of the expert knowledge. In the literature this is known as a

bottleneck problem of knowledge acquisition from an expert [Durkin 1991, Waterman 1989]. One of the suggested solutions is application of statistical, or machine learning based methods for the automated acquisition of knowledge from examples. It is assumed that it is easier to ask the user to provide examples from his past. In this approach the initial profile is created on the basis of responses to specific questions focused on her past travel experience. For example the system may ask user the about her last holidays:

- were you satisfied with your last holiday trip?
- and then either change the topic (if the answer is negative) or continue:
- where did you spent it?
- how did you arrived there?
- with whom did you spend it?
- etc.

Next the expert system analyses the answers and uses them to create the user profile. This way of gathering information may be very valuable, not only to build initial user profile. Applying cluster analysis we can create new customers classes and/or discover new significant characteristics of the user. This information may be also used by the associate rule techniques to discover the connection between travel service consumption (e.g. people who rent cabriolets are interested in dance clubs). The main problem with this approach is that the new user is asked to participate in an Eliza-like dialogue, which he may find confusing and thus unpleasant (why am I supposed to talk about all of this?). Additionally, he may find some of the questions invading his privacy, while the dialogue-based approach does not make it easy and/or natural to avoid answering the questions.

4. Initialization by observing the user's behavior

This is the user-friendliest approach, in a sense that it doesn't require the potential customer to answer any questions. It is also the approach that in many ways invades her privacy the most. Here, the system observes the user behavior on the Internet and collects data about her activity in the user behavior database. The system can potentially collect the data also about visitors who do not register in the system. Technology similar to cookies can be applied (a cookie is send to her computer during the "incidental" visit). When the visitor returns to become a customer, the collected information can be used to create her profile [Rud 2001]. There are a few problems with this approach. When the user enters the system and immediately becomes a customer, there may not be any data available to generate the profile (assume, for instance, that this user has just completely re-installed the operating system and our travel system is the first place she happens to visit). Additionally, this approach, while does not require the user to fill-in questionnaires intrudes into her privacy, and is therefore an example of problem (6) from Section 2.2. The advantage of this approach is that it can also be applied to the existing customers. Data concerning their Internet behavior can be collected and utilized to improve the accuracy of their profile as well as to discover new rules for the *specialized expert systems*.

Summarizing, there exists a number of techniques to initialize the user profile, but each one of them has significant disadvantages. Our initial analysis indicates that one will have to find a balance between creating an inaccurate initial profile, boring potential customers by long questionnaires, involving them in a confusing and/or uncomfortable dialogue, or invading their privacy. Here, one can almost hypothesize that different people would react better and would be more willing to work with different of the four approaches. Until we can find a new solution that overcomes the difficulties described above, the correct way may be to implement all four approaches and let the user pick the one that they feel most comfortable with (and making this determination will then be one more function that will be incorporated into the *user profile initialization expert system*).

5.3 Non-standard customers

One of the bigger problems for any *PIS* is the ability to deal with unconventional users (problem (2) from Section 2.2.). Observe first, that in a system that has correctly defined user-profile-clusters, it will be relatively easy to spot an "outcast." His profile will be far removed from any of the existing clusters. After such a customer is recognized we can apply a number of techniques that should be able at least partially cover-up the problem and keep the customer. Note that in the case of the travel advisory system, it is OK to provide a collection of possible choices out of which some are complete misses (as long as at least one fits the user). When interacting with non-standard users we will therefore mix suggestions based on: "near-by" profile clusters, "far-removed" profile clusters and random suggestions. If we are lucky enough (as clearly, in the case of eccentric customers luck is what the merchant needs), we will be able to keep the user with us long enough to be able to fine-tune his personal profile to be able to correctly personalize his content.

5.4 Delivery of information to the existing customers

We will now discuss issues related to the delivery of personalized information to the existing customers. We are particularly interested in problems (3) and (4) from Section 2.2. There are three situations that need to be addressed: (a) slow change in customer interest, (b) abrupt change, (c) overpersonalization.

The simplest problem to deal with is the slow change in customer' interests and we should be able to address it adequately. A typical example of such a situation is when over time customers' financial situation is improving. As a result he starts to use more expensive hotels, or rents larger cars, or goes to more up-scale restaurants. Another example would be a customer who stops accepting accommodations in Ramada's (maybe she did not like the service the last time she visited it and decides to not to stay in them anymore). Similarly, client may suddenly start requesting accommodations in IBIS hotels (her secretary made a mistake and made an arrangement in one and she liked the service). While these facts can be considered an abrupt change, observe that they are only changes in one of the characteristics stored in the user profile, while its remaining parts stay unchanged. We need to keep in mind that all facts about user behavior are recorded in the *user behavior database*. As soon as this change of behavior becomes pronounced enough for the appropriate expert system to recognize it, the user profile will be automatically updated (since we usually make more than one suggestion, a potential inaccuracy in the system should not be significant). Observed change of preferences may be also a part of the new trend and our customer is one of its "leaders." The expert system that is mining the information stored in the *customer behavior database* will be responsible for recognizing such a situation (e.g. by noticing the fact that multiple customers start renting hybrid power, or electric cars). In this case not only individual profiles need to be adjusted, but also the rules in other expert systems (e.g. users with profiles similar to these who have requested electric powered cars should be offered them as a choice as well) and possibly the user profiles may need to be re-clustered.

The situation is slightly more complicated in the case of an abrupt change in interests. Here, we do not have data that would allow us to spot the trend and make adjustments. Such a situation will manifest itself with the customer rejecting most/all of the suggestions provided by the system. This type of customer behavior starts to resemble the case of the non-standard customer (Section 5.3) and similar techniques should be used. We can, for instance, make suggestions based on: random selection, special offers and extreme suggestions (suggestions that are far-removed from these fitting the current customer profile). If any of our selections will be accepted we should use them to try to find which (different) customer profile would match these preferences well, and try to make additional selections based on this profile. We should also note in the customer profile that sudden change has been observed and during the next encounter apply this knowledge to extend the scope of suggestions. Future interactions with the customer should help modify the user profile in a similar way as in the case of slow changes.

Finally, to prevent the overpersonalization, we have to keep providing customers with information outside of the scope of their usual profile. Again, a mixture of special (seasonal) offers, extreme selections and random selections should be used.

5.5 Influences from the outside

One of the most difficult problems to dealing with is the influence of special events (problem (5) in Section 2.2.). Some of them can be predicted and included into our system (e.g. the fact that every four years large numbers of tourists will attempt to visit sites of Winter or Summer Olympics, or that Fall Comdex takes place in Las Vegas every November and attracts a particular group of system users, i.e. IT professionals). Others cannot be easily predicted (e.g. military coup in Malaysia, resulting in a civil war, or an outbreak of foot and mouth disease in Europe, that result in a complete change in travel patterns of our users as well an event cancellation etc.). Unfortunately, we have to admit that we do not have an easy way of dealing with this type of outside influences.

5.6 Privacy of data

The question of information privacy is of extreme importance in the case of development of the *PIS*. Here, the quality of the system depends on how closely is the customer willing to work with the system and how much of her personal information is she ready to provide. The ideal situation would be if the user was wiling to share with the system all of her personal data and allow the system to track her Internet behavior. Combination of this data would allow the system to fine tune her profile. To come anywhere close to this goal, we have to make sure that the customer understands that no personal information will be released to the third parties. In addition, we have to make sure that users believe that the system is safe, and thus their personal data that is being collected is not vulnerable to hackers (or the government).

5.7 Customer flow

It is clear that keeping an existing customer is much more efficient that attempting at finding his replacement

in the case when he drops form the system (we already have at least the initial profile to work with). At the same time, the system has to have procedures to deal with the data left by users who decided to abandon our system ("stale" data). Regardless of how well our system works, some customers will register and not use it, while others will use the system only a few times and then disappear. There are at least three ways of looking at the data left behind. On the one hand, all information is valuable as it adds to the total amount of accumulated knowledge. This being the case, especially in the early stages of the system life (when only a few customers are present), we may want to keep all data. On the other hand, initial profiles that have not been properly tuned and short sequences of customer behavior can be treated as garbage as they represent the information that is least trustworthy e.g. in the rule mining process. Therefore, as the system matures we may want to clean our databases from the stale data. Note that after a large amount of incidental data is removed a process of re-tuning the system may need to be applied to make sure that the rules that exist match the remaining data. It is possible that some of the rules that have been mined from the user behavior database could not have been extracted from the remaining (cleaned-up) data and therefore may need to be removed, or at least their importance adjusted. Finally, one may want to analyze the information available in the data associated with incidental customers. For instance, this may help us discover the reasons for their stopping using the system and help improve its overall friendliness.

6. Concluding remarks

The aim of this paper was to introduce the main issues involved in creation of a travel oriented personalized user information delivery system. We have indicated seven major problems involved in creation of *PIS* and attempted at addressing them in the context of our earlier proposal for an agent based travel support system. We have sketched the general structure of the *PIS* subsystem, consisting of the *user interface*, a collection of *expert systems* and two databases (*user profile database* and *user behavior database*) and presented interactions between them. Here, we have concentrated on the general structure of the system and an attempt at addressing the top-level problems in the *PIS* development. There remain a number of important questions that we will attempt at addressing in the near future.

Interacting with the experts from the travel industry we will have to analyze the market to define the characteristic features that define the customer and will be a part of the user profile as well as the ontology of travel services [Grunninger 2001]. These will be used, among others, to develop the questionnaires and become a part of the expert systems. We will have to define the types of expert systems that match the functionality of each of the individual expert systems defined in Section 5.1. We will have to analyze the appropriate data mining algorithms that are to be used to the *user profile database* and the *user behavior database* to discover the appropriate types of knowledge to be used in the system. Real-time aspects of the system will have to be addressed (e.g. to address questions: how often to attempt at modifying the user profiles? how often to attempt at finding new rules for the individual expert systems? etc.). Questions related to dealing the data resulting from the incidental users (both the issue of their not using the system and what to do with the data after they leave). Finally, a set of questions related to the business model has to be addressed. We plan to start addressing these and other points left open here in the near future.

At this time we have initialized the system implementation process. Based on the analysis presented in [Rahimi 2001] we have selected the Grasshopper agent platform and the JESS expert system environment. Our initial goal is to finalize the agent decomposition of the system and implement all of the necessary agent classes and their interactions.

7. References

[Abramowicz 2002] W. Abramowicz, P. J. Kalczyński, "Building and Taking Advantages of the Digital Library for the Organizational Data Warehouse," in: M. Cobb et. al. (eds.), *Proceedings of the Second Southern Conference on Computing*, Hattiesburg, Mississippi, USA, October 26-28, 2000, to appear

[Brady 2002] R. Brady, E. Forrest, R. Mizerski, *Marketing w Internecie*, PWE, Warszawa, Poland

[Brenner 1998] W. Brenner, R. Zarnekow, H. Wittig, C. Schubert, *Intelligent Software Agents*, Springer-Verlag, Berlin, Germany

[Chaudhury 2001] A. Chaudhury, D. N. Malik, H.R. Rao, "Web Channels in E-commerce," *CACM*, 44(1), pp. 99-104

[de Kare-Silver 2001] M. de Kare-Silver, *E-shock: the New Rules. E-Strategies for Retailers and Manufacturers*, Palgrave, Houndmills, UK

[Dreyfuss 2002] J. Dreyfuss, "Blocking the Pitch," *Mobile Computing and Communications*, March 2002, p. 40

[Durkin 1991] J. Durkin, "Designing an Induction Expert System," *AI Expert*, 12, pp. 29-35

[GIS 2001] Open GIS Consortium, *Call for Participation in the Open Location Services Testbed Phase 1*, Bloomington, Indiana

[Grover 2001] V. Grover, J.T.C. Teng, "E-commerce and the Information Market," *CACM*, 44(4), pp.79-86

[Grunninger 2001] M. Grunninger, J. Lee (eds.), "Ontology, Applications and Design," Special Section of the *CACM*, 45(2), pp. 39-65

[Jennings 2001] N. R. Jennings, "An agent-based approach for building complex software systems," CACM, 44 (4), pp. 35-41

[Lieberman 2001] H. Lieberman, C. Fry, L. Weitzman, "Exploring the Web with Reconnaissance Agents," *CACM*, 44(8), pp. 69-75

[Lohse 1998] G. L. Lohse, P. Spiller, "Electronic Shopping," *CACM*, 41(7), pp. 81-88

[Manber 2000] U. Manber, A. Patel, J. Robinson, "Experience with Personalization in Yahoo!" *CACM*, 43(8), pp. 35- 39

[Mohr 2001] J. Mohr, *Marketing of High-Technology Products and Innovations*, Prentice-Hall, Upper Saddle River, NY

[Ndumu 1998] D. Ndumu, J. Collins, H. Nwana, "Towards Desktop Personal Travel Agents," *BT Technological Journal*, 16 (3), pp. 69-78

[Paprzycki 2001a] M. Paprzycki, R. Angryk, K. Kołodziej, I. Fiedorowicz, M. Cobb, D. Ali, S. Rahimi, "Development of a Travel Support System Based on Intelligent Agent Technology," in: S. Niwiński (ed.), *Proceedings of the PIONIER 2001 Conference*, Technical University of Poznań Press, Poznań, Poland, pp. 243-255

[Paprzycki 2001b] M. Paprzycki, P. J. Kalczyński, I. Fiedorowicz, W. Abramowicz, M. Cobb, "Personalized Traveler Information System," in: Kubiak B. F. and Korowicki A. (eds.), *Proceedings of the 5th International Conference Human-Computer Interaction*, Akwila Press, Gdańsk, Poland, pp. 445-456

[Paralic 2001] J. Paralič, M. Paralič, M. Mach (2001) "Support of knowledge management in distributed environment," Informatica, 25(4), pp. 319-328

[Rahimi 2001] S. Rahimi, R. Angryk, J. Bjursell, M. Paprzycki, D. Ali, M. Cobb, M. Gibutowski, K. Kołodziej, "Comparison of Mobile Agent Frameworks for Distributed Geospatial Data Integration," in: Konečny M. (ed.), Proceedings of the 4th AGILE Conference on Geographic Information Science, AGILE, Brno, pp. 643-655

[Rud 2001] O. P. Rud, *Data Mining Cookbook. Modeling Data for Marketing, Risk, and Customer Relationship Management*, Wiley, New York, NY

[Strauss 2001] J. Strauss, R. Frost, *E-Marketing*, Prentice-Hall, Upper Saddle River, NY

[Waterman 1989] D. Waterman, *A Guide to Expert Systems*, Addison Wesley, Boston, MA