PERSONALIZED TRAVELER INFORMATION SYSTEM

Paweł Jan Kalczyński, Marcin Paprzycki, Iwona Fiedorowicz[1],

Witold Abramowicz, Maria Cobb

Abstract: In this paper we propose a model of a Personalized Traveler Information System (PTIS) that brings together functionalities of mobile Internet, Geographical Information Systems (GIS) and information filtering. Our idea is based on the fusion of travel information with geographic localization, and on the use of information filtering techniques to provide users with the relevant data. Agent system that assists users searches pre-selected Web resources, on the basis of localization and preferences, to find the suitable information. In this way users are provided with more precise information (in comparison with that available among currently existing Internet services). The proposed system is based on already-available solutions and Internet standards.

1. Introduction

Personalized Internet Services have become an important topic of research and real-life applications. Almost all Internet Portals offer a personalized view of their resources. Some of them go beyond naive profiles and offer a truly personalized information delivery [1]. At the same time, personalization seems to be somewhat neglected in Web-based travel information services[2], where it typically allows for a few simple preferences, e.g., related to hotel rooms and airplane seat assignments [2]. In addition, we should note that while there exists a large number of repositories containing travel-related information, these services are not integrated. For instance, the restaurant information may be stored in one repository, while the hours of operation of a museum will be available in a different one. Let us also note that even in the case of integrated services, they may integrate only selected businesses while omitting others. For instance, some time ago Southwest Airlines elected not to pay a fee to the SABRE reservation system (one of the largest airline reservation systems in the world) and as a result, it was not included in any of the SABRE-based travel services. Similarly, it is often the case that only the more expensive accommodations are available via the Internet travel agencies, while the cheaper, non-chain motels and B&B's are omitted. It is our belief that only a system combining data stored in multiple repositories can provide the traveler with high quality information (for more details on delivery of precise personalized information see [3]). In order to be able to properly address the needs of the traveler, such a system should combine mobile Internet, Geographic Information Systems and Selective Dissemination of Information systems.

The concept of selective distribution of information (SDI) was derived from the concept of selective dissemination of information created by H. P. Luhn [4] in order to

---

[1] Corresponding Author

[2] Note that 'travel' has a broader meaning than 'tourism'. For instance we may think of business travelers, tourists and local-pub-visitors as users of Personalized Travel Information Systems.

improve scholarly communication among universities. SDI systems should provide users only with relevant information, while rejecting irrelevant data. In the SDI system, information producers supply digitally represented source information, which is then indexed and used for fast delivery of information precisely matching users' needs. The idea of applying agent-based SDI systems to directly supply information systems (instead of humans) with relevant documents was introduced in [5] and developed in [6] [7] [8] [9] with the research focused on building the knowledge base for the information system (server side of the system). Separately, in [10] we have discussed the agent-based infrastructure for the development of the travel support system (which can be considered as the middle layer of the system). In this note, we will concentrate our attention on the client-side of the system (functionality and interface).

2. General Architecture of the Proposed System

The general architecture of our proposed solution is illustrated in Figure 1, and its basic elements are: content providers, data federations (developed on the basis of the Geographical Information System), agent-based Selective Dissemination of Information system and Internet devices (static and mobile). Due to the space limitations, we will provide only a rough description of each of the system components.

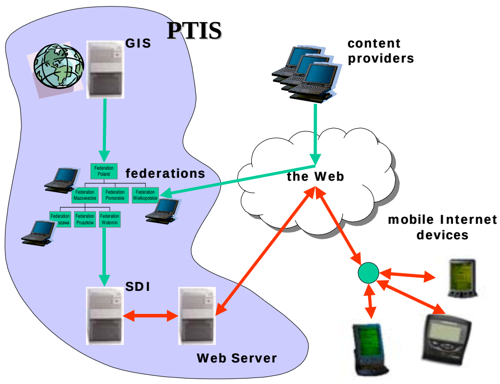

Fig. 1. An overview of the Personalized Traveler Information system

2.1. Federations and GIS

In the proposed approach, the information provided by the content providers is *indexed* according to the geographical location. An object that represents a particular geographic region will be referred to as a **federation**. Federations are the way of combining the geographical information from the GIS system with other travel-related information. The initial division into federations is likely to proceed according to the administrative division, typical for a given country (e.g. country, state, county). On the lower levels, it will proceed according to the rules established by the last-level-federation supervisors. Alternatively, other geographical division schemas may be applied (e.g. religious, historical, economical). In this way, the geographical information represented in one of the standard geo-representations, e.g. a quad-tree, is translated into a travel-oriented hierarchical representation. Let us note that, while federations are organized into hierarchies, federations that belong to the same level of a hierarchy cannot overlap. The size of the federation may range, for instance, from the continent to the city museum. While using the system, users may chose federations, and switch between the schemas. The schema depicted in Figure 2 illustrates the hypothetical section of the division of Poland.

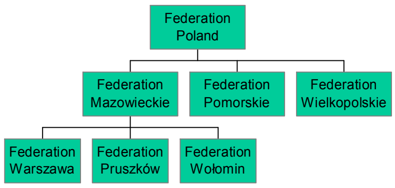

Fig. 2. An example division of the information into federations

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 assigned to appropriate federations. 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 (see Figure 2). In order to be able to achieve this, documents have to be described by metainformation, which will serve as the input for the GIS and SDI systems. Metainformation can be encoded in one of the standard content descriptions, e.g. RDF (Resource Description Framework) or simply in the META section of HTML documents. As an example consider the metainformation record given below:

```
<META name="provider_id" content="1235"
<META name="federation_id" content="2.17.13.124.246677">
<META name="doc_type" content="transport.bus.timetable">
<META name="valid_thru" content="2001-10-11">
```

The structure of the document type in the above example must be standardized beforehand. The indexing function (assigning document metadata to the correct federation{s}) will be carried out by *indexing agents* [8]. In case of a single federation, hierarchy (e.g. based on the administrative division) a document will be assigned to the lowest level federation that completely encloses its geographical extent. In the case of multiple federation hierarchies (e.g. administrative + historical + religious etc.) a given document may be indexed in all of them (depending on its content; e.g. bus schedule may, or may not be included in the history-based hierarchy of federations).

Observe, that we are building the system around an "indexing hierarchy" storing information about the location of documents containing pertinent information. This approach to the development of WEB oriented information system has important advantages (for more details see [8]). First, only a limited amount of information is stored on the server. Second, the information is already ordered and pre-processed which allows a more efficient organization of the search process (also see Section 3.3 below). Third, limited amount of information to be searched along with appropriate organization of documents positively affects the efficiency of the retrieval process (in terms of *recall* and *precision* measures). When the processing time is considered, since only a subset of potential documents is selected to be relevant to a given search, two outcomes are possible: (1) in a given time a more thorough processing of selected data is possible, or (2) answers can be found substantially faster. Considering search precision, according to its definition, retrieval / filtering precision is the number of relevant documents divided by the number of all documents retrieved by the system as the result of processing a given query (profile). The total number of to-be-searched documents is limited to the selected federation, its parents, children and selected siblings. The overall probability of retrieving irrelevant documents is diminished, as documents extremely likely to be irrelevant to a given search are not processed. Hence, the proposed division of the globe into federations may be conducive to more precise travel information filtering and retrieval.

2.2. Filtering Agents (SDI System)

Personal Assistant Agent is a software agent, which acts on behalf of the user in the Selective Dissemination of Information system. In order to fulfill its task, the agent requires information about user requirements/needs. This information will be derived from three sources. First, an expert system containing rules matching the types of travel ,e.g., business, family vacation, romantic getaway with activities typically associated with them (e.g. business dinner, visit to an amusement park etc.). This expert system will be originally created by travel industry experts, but it will evolve over time. There are two ways in which this will happen: (1) analysis of the behavior of all users of the system will lead to the rule modification, (2) for each user, a set of individualized rules will be developed (a personal extension to the expert system). The set of individual standard travel preferences (e.g. aisle seat, non-smoking room, economy car, Thai food etc.) will originate in the form of a questionnaire filled-in when the user account is initialized. As in the case of the expert system, the SDI can extract and fine-tune the information about the preferences based on the actual choices (system learns about the user). Finally, the set of current preferences may differ, or be more detailed than the standard set (e.g. when flying to Seattle I may want to have a window seat to see Mt. Rainier). Let us also note that information about: (a) number and age of people on the trip, (b) gender, (c) age, (d) nationality, (e) ability to speak foreign languages, (f) ability to change location (mobility) etc., is likely to find its way to all three levels of travel profiles. Profile sets should be encoded in the way that enables serialization (e.g. in the XML compatible language). Depending on the available information, the SDI distinguishes services that are useful for users during a given trip (for more details see Sections 3.1 and 3.2).

2.3. Content Providers

Today, there exists a very large number of WEB-based providers of information that can be useful for travelers. In addition, Internet portals, containing large sets of travel-related information, are also being continuously developed. Useful travel-related information may be grouped into (this list is definitely not exhaustive):

- weather information: [11] [12] [13]
- information about dangers: [14] [13]
- maps: [15] [16] [17] [18]
- sites of interests – most vortals
- accommodation: [19] [20] [21] [16] [17] [18] [22] [23] [24]

- dining: [16] [17] [18] [22] [23]
- timetables and fares: [17] [23] [18]

We will skip over the question of selecting "trustworthy" content providers, as well as the tasks of information validation and dealing with conflicting information, as these questions, while definitely of great interest, are outside of the scope of the current paper. Let us assume that we were able to identify a number of trusted content providers. As specified above, they will deliver meta-information that will be indexed into appropriate federations. It should be noted that in an agent-based system it is possible to augment the trusted information sources by general Internet searches quite easily [25] (performed as low-priority activities); however, this area is also outside the scope of this paper.

2.4. Mobile Internet and the GPS

In addition to standard PC-based browsers, there are already many ways of using mobile Internet: from the, very common in Europe, Wireless Application Protocol (WAP) browsers – through laptop computers (linked via mobile phone modems) – to palmtop units with built-in mobile modems. As advantages of mobile Internet are commonly known, we will limit ourselves to listing a few of its drawbacks (in contrast to the desktop PCs). Slow links, limited display capabilities and short life of batteries are the main problems for the mobile Internet. It should be also pointed out that even the smallest devices are often too big, too fragile or not user-friendly enough to play the role of personal travel assistants. Nevertheless, the mobile Internet is evolving and we believe the 3G devices may provide solutions to these problems. In addition, agent based systems, as those described in [25] can help to overcome some of the difficulties faced currently by mobile Internet users.

The GPS system was originally available only for the army. Later, it became available in sailing and aviation. Recently, GPS systems are also available, among others, in luxury cars (e.g. Street Pilot) or for advanced trekkers (e.g. eTrex), fishers (e.g. Fishfinder) and sea-sailors. The constellation of 24 satellites, orbiting 10200 km high, allows specifying the worldwide location (in three dimensions), with precision varying from one to several seconds of longitude and latitude. While there exists a large number of questions related to the privacy issues raised by this technology (especially in the context of the legislation proposed in the US that would require all cell phones to include GPS traceable devices), usefulness of the GPS in the personalized traveler information system is obvious. A GPS system establishes location of a user, which allows the system to provide the relevant information. We can expect that in the near future a combination of GPS and mobile Internet and our system is capable of taking advantage of such functionality.

3. Functionality of the System

3.1. User Accounts

Users have their own accounts in the system, which enable them to specify their current travel needs, define and modify profiles, and manage information acquired by the system. When opening an account, a user fills in a questionnaire describing her general preferences (see also Section 2.2). This information will be then combined with the general rules matching travel type with expected activities and with the information about the current trip to search for the relevant information. Since the functionality of the system resides in the server, users can access their accounts through the traditional Web browsers, as well as mobile devices. Thus, ATM-like Web terminals that enable users to access their PTIS accounts, might substitute traditional tourist information centers. Some airports (e.g. New Orleans, Dallas or Chicago) and cities (e.g. Zakopane) already have Web terminals available to public. Regarding the efficiency of the system and the connection speed, user accounts should migrate among servers (it is assumed that a fully developed system will have to rely on a distributed server farm rather than a single server), depending on the geographical location. Information about user location will pass the signal to the closest (in terms of the connection speed) mirror server and initiate the automatic migration process. At the beginning of a new trip, a user logs-in to the system and provides it with the travel details as well as a set of current preferences. This action triggers the SDI system to start searching for the relevant data. Since the system is based on the agent technology, as soon as the search process is initiated, a user can log-out and come back at some later time and access the results.

3.2. SDI System

To support a given trip, a travel assistant that works as a part of the Selective Dissemination of Information system, combines the three sets of data: travel expert data, users' general preferences and users' current preferences (with current preferences superseding the general ones in context of a given trip). This information is used to selects terms (words and key phrases), which will be used to search for the travel information. The agent filters active sources of travel information and retrieves information from the passive ones (registered in the appropriate federations). Documents that are compatible with user contexts (location, profile, time, destination) are presented to the users via their accounts. Hence, the filtering process consists of:

(1) generating a subset of federations that may contain information the user is interested in,

(2) for all selected federations, comparing metainformation with the profiles and selecting a set of documents to be searched,

(3) performing the traditional keyword-based information retrieval query on the textual parts of documents selected in (2),

(4) cross-referencing and validating the information retrieved from multiple sources, and

(5) preparing the filtered information for display.

When the subset of the federation hierarchy is generated, it includes "parents" of the federation that the user is currently associated with (for federation Pruszków, metadata stored in federations Mazowieckie and Polska will also be searched; see Figure 2). In addition, all "children" of a given federation are also traversed in search for pertinent metadata. The search may include “siblings” of a given federation. Here, however, the search depends on the predicted mobility of the traveler. For instance, federation Pruszków will be included in the search process only if it can be predicted that the traveler, who is currently associated with federation Warszawa will travel to Pruszków. The same schema will be applied in the case when multiple federation hierarchies are to be searched for pertinent information (e.g. for someone interested in historical churches federations based on administrative divisions as well as religious ones have to be traversed).

We will assume that research issues related to data cross-referencing and deconfliction, as well as validation of the obtained results, have been addressed successfully and the system has these capabilities built-in. Let us focus our attention on the question of output generated by the system. On the basis of active data sources (e.g. flight delay info or weather advisories) users may be notified of the high-priority information (event-alerting). The system has to provide a selection of data display capabilities. First, users should be able to access all collected documents. In this case, the final result of the search is a number of files that are returned to the server and made available to the user. These documents and references should be presented to the user through the personalized browser-based interface in a hierarchical layout (e.g. similar to e-mail folders). Obviously, this option is desirable only if the user has appropriately fast Internet access and a reasonable amount of available time. Second, collected documents are further filtered to prepare a summary of results (with the filtering process taking place on the server, or remotely via agent collaboration). Here, the level of condensation of the results depends on at least on two factors. When the limits are imposed by the parameters of the access device (or – more precisely – the browser type), the Web Server should adjust the presentation of information to the appropriate interface (final level of information filtering), such as a typical browser, Windows CE browser, Palm OS browser or WAP browser. In case of the person-related limitations (e.g. severe shortage of time, or reduced attention span), the level of result processing (filtration and condensation) should be a menu driven option. As with other functionalities listed above, we expect that the system will start with predefined “default” characteristics. Over time, based on the interaction with the user, the system will fine-tune/personalize the interface.

3.3. Using Federations to Rank Retrieval Results

Federational information filtering and retrieval enables the application of assumptions of the hierarchical model to rank the results of query (profile) processing. For instance, during the excursion users will most probably prefer documents, which precisely describe their context. In other words, they will prefer documents assigned to the federation that is placed on the lowest possible level of a given hierarchy. On the other hand, users who only plan their journeys may be equally interested in documents assigned to upper-level federation and its peers. Since federations are hierarchical, user preferences may be easily translated into a normalized modifier $\mathbf{f_m}$. The modifier may then be applied to change the ranking of information filtering / retrieval results. The example modifier $\mathbf{f_m}$ is given by the formula below.

$$f_m = \frac{1}{(1+u)^d}$$

where, $\mathbf{f_m} \in [0;1]$ – federation modifier of ranking of filtering / retrieval results, **u** – level upwards of federation (relative to the one a user is currently in), $\mathbf{d} \in [0;5]$ – preferred detail level: 0 – all federations are equally important, 5 – documents from the current federation are the most important. In this context, assuming that d=0 means that documents from all upper-level federations (u=1,2,3,..,n) will be considered as equally important to the user. Whereas, for d=2, documents from the parent federation (u=1) will be four times less important than documents from the current federation (u=0). Obviously, initially there will be a set of default weights assigned to all accounts. Over time these weights will be personalized for each user (based on the user interaction with the system).

3.4. Cross-Federation Filtering and Retrieval

The division of a particular area into separate federations should not restrict the functionality of the system. Some geographical objects, such as rivers, roads or mountain ranges may belong to many lower-level federations. The GIS task is to handle and resolve such assignments into the proper division. Given a detailed location data, GIS finds federations that virtually constitute a given geographical object (windowing). Note that means of transportation (specified in the profile) may determine the radius, within which federations are searched. Such a solution is applied in some virtual maps (e.g. attractions search in Microsoft Auto Route).

4. Possible Advantages and Disadvantages of the Proposed Approach

4.1. User-side Advantages and Disadvantages

The advantages of the proposed system include: separation of the user from the physical device (search for answers takes place while the user is off-line), shortening of the time needed to search for the travel information and/or a higher – in comparison with traditional models – precision of information filtering and retrieval.

The main drawback of the proposed system is the necessity of connecting to the server. This may be particularly difficult or expensive (prices of satellite phone connections) during e.g. sea sailing or mountain trekking. Other disadvantages may include typical problems with the Internet information services such as information overload and out-of-date information. One may hope that carefully selected content providers (high-quality information) may help to reduce these difficulties.

4.2. Provider-side Advantages and Disadvantages

The main advantages of the proposed system from its provider's point of view may include: new possibilities of analysis of tourist behavior and availability of high quality marketing data allowing for personalization of advertising. Sophisticated analyses performed on user profiles and tracking user activities in the system (log analysis) may result in increasing the knowledge on the tourist services market. This knowledge may be particularly useful for travel agencies, hotels, restaurants and others who are somehow connected with tourism. What is more, this knowledge enables the provider to specify a marketing target more precisely. This may make the system attractive for advertising.

The biggest challenges for the system are: the need to handle heterogeneous resources, large number of processes and large volume of documents (high CPU and storage requirements) and the need for frequent updates and synchronization of information collected from distributed sources (for problems involved in the development of a system that deals with only heterogeneous geospatial integration see [10]). The key point here is the need of defining the rules that assign documents to selected federations. Proper definitions of such rules will determine the functionality of the whole system.

5. Summary and Future Work

In this paper, we have sketched a model of the personalized traveler information system and highlighted its key points. The ideas presented here resulted from our earlier research and experiments with information filtering and retrieval from heterogeneous sources on the Web. We realize that building a single global solution based on the proposed model is economically irrational and technically nearly impossible. However, the proposed solutions might be applied, tested and improved locally. Almost every city has its information portals, which could be personalized in the way we suggested. The experience gained during creation of the first municipal portal might help to build other tourist information services. Such personalized portals (e.g. Warsaw, Poznan, Bydgoszcz) might be combined into a regional federation and – finally – into national federation (Poland). Therefore, the system may be developed bottom-up. However, in order to implement the proposed model, a set of standards must be supplied and the most important are: division into federations and content description framework. We believe that the constant growth of the tourist services market will force the service providers to come up with such standards.

6. References

[1] *My Fidelity* www.fidelity.com

[2] *Travelocity* www.travelocity.com

[3] W. Abramowicz, P.J. Kalczyński, K. Węcel *Profiling the Data Warehouse for Information Filtering* Managing Information Technology in a Global Economy (M. Khosrowpour ed.) Idea Group Publishing, USA, 2001, 810-814

[4] Luhn H.P., *A Business Intelligence System*, IBM Journal of Research and Development, Vol. 2, No. 2, s. 159-165

[5] W. Abramowicz, P.J. Kalczyński, K. Węcel *Information Filters Supplying Data Warehouses with Benchmarking Information* Chapter 1 in Knowledge Discovery

for Business Information Systems, W. Abramowicz & J. Zurada (eds.), Kluwer Academic Publishers, Boston, 2001

[6] P.J. Kalczyński *Supplying Data Warehouses with Benchmarking Information* Proceedings of Lubniewice 2000 (May 2000), Economy In Transition (A. Janc ed.), The Poznan University of Economics Publishing House, Poznań 2001

[7] W. Abramowicz, P.J. Kalczyński *Intelligent Agents to Supply the HyperSDI System* Technologie inform. w zarzadz. - SWD (J.Studziński, L.Drelichowski, O.Hryniewicz, J.Kacprzyk eds.), IBS-PAN, Warszawa 2000

[8] W. Abramowicz, P.J. Kalczyński *Building and Taking Advantages of the Digital Library for the Organizational Data Warehouse* Proceedings of the Second Southern Conference on Computing, Hattiesburg, Mississippi, USA, 26-28 Oct. 2000

[9] W. Abramowicz, P.J. Kalczyński, K. Węcel *Time Consistency among Structured and Unstructured Contents in the Data Warehouse* Managing Information Technology in a Global Economy (M. Khosrowpour ed.) Idea Group Publishing, USA, 2001, pp 815-818

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

[11] weather.yahoo.com

[12] *Mathematical Weather Forecast* meteo.icm.edu.pl

[13] *Dangers and Weather in Polish Tatras* www.cyf-kr.edu.pl/tpn/kom/aktualny.html

[14] *Worldwide tourist information on dangers* www.fieldingtravel.com/df/index.htm

[15] Virtual USA and Canada maps.yahoo.com

[16] *Virtual Tourist* www.vtourist.com

[17] *Italian Tourist Web Guide* www.itwg.com

[18] *Search NYC – Cyber Guide of New York City* searchnyc.com/tourist.html

[19] *Polish Tourist Portal* www.turystyka.net

[20] *Polish hotels* www.hotelepolski.com.pl

[21] *Accommodation in Poland* www.turysta.com.pl

[22] *Tourist Guide of Greece* www.odysseas.com/index.html

[23] *Poznań Municipal Portal* www.info.poznan.pl

[24] *Zakopane On-line Information Center* www.zakopane.pl

[25] K. Kołodziej, M. Paprzycki, M. Cobb, D. Ali, *Intelligent Agents in a Distributed GIS for Improving Interoperability*, Proceedings of the PIONIER 2001 Conference, (S. Niwiński ed.), Technical University of Poznań Press, Poznań, Poland, 257-266