THE EVALUATION OF A RECOMMENDATION SYSTEM FOR TOURIST DESTINATION DECISION MAKING

Ulrike Bauernfeind
Institute for Tourism and Leisure Studies
Vienna University of Economics and Business Administration
Augasse 2-6
1090 Wien
Austria
Ulrike.Bauernfeind@wu-wien.ac.at
ABSTRACT

The present paper will outline the creation and evaluation of an intelligent tourist recommendation system named DieToRecs. Recommender systems became a significant tool in the field of tourism; they offer users a convenient opportunity to find a travel bundle or a single travel item such as accommodation. Existing tourist recommendation systems have some shortcomings as they allow no or only very limited flexibility when taking constraints or preferences into account. Therefore, the aim to design a novel type of recommendation system including features such as Collaborative Filtering and Case Based Reasoning (CBR) is pursued by the DieToRecs project team. Before the final system will be implemented, two prototypes are created. Experts and user groups evaluate each of the prototypes to assure an improvement and to enable the best final version possible. The evaluation considers aspects such as design and layout, functionality or ease of use. An aim is to analyze the results of the assessments. The ultimate goal of the thesis is to realize a meta-evaluation meaning to assess which methods are the most effective and useful ones regarding the analysis of tourist recommendation systems.

1. INTRODUCTION

The present paper will outline the creation and evaluation of a tourist recommendation system named DieToRecs. The project is funded by the EU (under contract DIETORECS IST-2000-29474 and within the IST programme) and aims to create an intelligent recommendation system for tourist destination decision making. Recommendations systems in the context of travel and tourism became increasingly important since the amount of available information is exploding and users are not always experienced in processing the multitude of information resources (Hwang, Gretzel and Fesenmaier 2002, Pühretmair, Rumetshofer and Schaumlechner, 2002). Thus, the need for a system supporting the user efficiently in travel decision making is obvious. Existing tourist recommendation systems have some shortcomings insofar as they allow no or only very limited flexibility when taking constraints or preferences into account. A practical example is that often too many results or on the contrary, no results are returned by the system (Ricci et al. 2002b). Current systems offer no support considering the reformulation of queries. Moreover, the comparison of different travel items is often not possible and no explanations for the recommendations are given. Further weaknesses are for instance that the opportunities to store information or to create user profiles are very limited.

The development of DieToRecs is an attempt to overcome these problems. The recommendation system will include a number of characteristics distinguishing this system from already existing ones. First, DieToRecs will be a destination advisory system based on CBR (Case Based Reasoning). A case is defined as any single user-system interaction history and Case Based Reasoning is a methodology trying to solve a problem by retrieving and using a similar, already solved case (http://dietorecs.istc.it/TechnicalApproach.html). A case consists of travel wishes and constraints, the travel plan (a bundle of items being interesting for the user and which is therefore collected), the user and the outcome (Ricci et al. 2002a). Arslan and Ricci (2002) added the navigation history to be a fifth component of a case. Second, a collaborative filtering system will be employed enabling recommendations related to similar users. User modeling and learning is another characteristic applied allowing the filtering of similarities of users. Finally, the system should be an interactive and conversational one meaning the user should feel like interacting with a human being. Queries and suggestions should follow successively to enable a vivid question and answer process. Interactive query management will be employed to handle queries more efficiently. The system will help the user to redefine queries; they are relaxed or tightened in order that a desired number of ranked results can be achieved (Ricci et. al. 2002b, Ricci, Arslan and Mirzadeh 2002).
The DieToRecs project started in July 2001 and will be completed by the end of 2003. There are five partners involved in the project: ecTRL, the Electronic Commerce and Tourism Research Laboratory (Italy), TIS, Tiscover Travel Information Systems (Austria), IFTL, the Institute for Tourism and Leisure Studies of the Vienna University of Economics and Business Administration (Austria), NLTeC, the National Laboratory for Tourism and eCommerce of the University of Illinois at Urbana-Champaign (USA) and finally, APT, the Azienda per la Promozione Turistica del Trentino (Italy).

A number of products and services, i.e. accommodation, attractions, accessibility, amenities and ancillary services can be recommended (http://dietorecs.itc.it/ProjectSummary.html). The system will enable the user to be advised in respect of two types of inquiries. First, single travel items, such as accommodation or destination can be recommended. Second, a suggestion of a complete travel selection, a bundle of several products, is possible as well. Figure 1 shows the start page of DieToRecs at present and the possibility to select between the recommendation of single items or a complete travel bundle is illustrated (in the main frame).

Furthermore, there are two main options involved in the recommendation procedure. The first one is that the user states important facts and preferences concerning his planned travel. The left frame of the web site displayed in Figure 1 shows this variant of recommendation. After the user has indicated his wishes, the system will then propose several items, i.e. destination, accommodation and activities according to these criteria. The results will have a ranking and will include an explanation of the ranking.

The second possibility called “idea pool” is that the traveler gets six templates of complete travel bundles. It is intended for users who have not yet decided which type of vacation and destination they are looking for. They might search for ideas and proposals. This option can be chosen in the main frame of Figure 1, it’s the last point called “Are you seeking for inspiration?” Figure 2 shows the site the user gets when choosing this option. On the basis of these proposals he/she can select which alternative/s he/she likes or dislikes and by means of these decisions new and refined options will be given.

---

**Fig. 1. The First Prototype of DieToRecs**

---

3
There are two ways to gain easy travel recommendations:

- Follow one of the alternatives you are interested in and you will receive detailed offers.
- Rate the alternatives, click "Submit" and you will receive additional alternatives.

Fig. 2. The “Idea Pool”

These two options do make sense, since there are different decision styles involved when interacting with a recommendation system (Grabler and Zins 2002). There are two approaches of information search: holistic (alternative-based) and analytical (attribute-based) strategies (Fesenmaier et.al. 2002). A system adapting to these two basic distinctions is likely to improve the human-computer interaction (Zins 2003). The DieToRecs recommendation system takes these findings into account by providing alternative-based and pictorial information in the so-called “Idea Pool”. This option corresponds as well to the recommendation-oriented decision style type identified by Grabler and Zins (2002) and follows the recommendation strategy proposed by them.

An additional amenity of the system is that the user will have the possibility to store the recommended item/s he/she regards as interesting in a so-called “travel plan”. Furthermore, the user will have the opportunity to create his own profile allowing him storing information concerning his preferences or an item he has chosen for his “travel plan”. This profile enables to recall the information in another session.

The project was basically decomposed into four temporal phases. First, a decision model has been developed to define the requirements and features the system has to have. The second phase consisted of the development of a first prototype. After the completion of this first version (in March 2003) an evaluation becomes necessary to assure an improvement and to provide a better second prototype. This assessment will be done in April and May 2003. The evaluation will focus on criteria such as functionality, ease of use or the design and layout. After the analysis of the evaluation will be completed, the first prototype will be improved. Additional features will be implemented to become the second prototype. Another assessment will be done to finally achieve the best last system version of DieToRecs.

One direction for research is the analysis of the evaluations regarding functionalities, design or outcome. Another useful examination is for instance to conduct user testing in both approaches - the recommendation versus the preference oriented approach (by control group strategy). Undertaking this confirmatory analysis could prove that both of the approaches are needed because of different user types and decision styles.
A further goal pursued by the study is a comparison of various methods recommendation systems are based on, such as search-based recommendations versus category-based or the application of collaborative filtering and clustering (De Bruyn 2001). Further objectives are to compare different assessment strategies and to show which methods are most effective and informative when evaluating a recommendation system.

2. THEORETICAL BACKGROUND

2.1 General Theories about Customer Satisfaction

First, some general theories about customer satisfaction and the assessment of services should serve to give an overview about what affects customer behavior. According to Trommsdorf (1998), there are a lot of factors influencing behavior such as involvement, knowledge, emotions, motives, attitudes and values, personality, the type of information acquiring and the kind of processing of information.

Vavra (1997) introduced a model of consumer satisfaction being dependent on expectations, perceived performance and prior experience. Prior experience is determined by a lot of antecedent influencers such as demographics, word of mouth, and nature of competition, advertising and expectations are formed by desires, perceived performance by the ease of evaluating.

Parasumaran et al. (1985) described a model of five dimensions of service quality: tangibles, reliability, responsiveness, assurance and empathy. Although the model was originally constructed for the traditional service sector it can be well adapted to computer-mediated environments, tangibles meaning the Graphical User Interface, reliability can be the error probability of the system itself or the trustworthiness of the recommendations given. Responsiveness can be related to the fact that a recommendation system should be a highly interactive and conversational one. Assurance is another aspect named by Parasumaran et al. which can be translated into the advisory system context to the extent of which a system is capable to give convenient recommendations, to have large knowledge (e.g. including many different types of destinations, accommodation) or that the system is able to convince the user about the goodness of the recommendations. Finally, empathy in the context of recommendation systems can be the degree to which the system is able to respond to the wishes and preferences stated by the user.

2.2 Theories about Human-Computer Interaction (HCI)

Second, theories, which concentrate on human computer-interaction and computer-mediated environments (CMEs), will be used to illustrate the factors, which mostly influence the usage of a system. The Technology Acceptance Model (Davis 1989) relies on two factors explaining human behavior: perceived usefulness and perceived ease of use. Perceived usefulness describes the user’s point of view of enhancing his/her performance by using the system. Perceived ease of use is the degree of effort the user believes he or she will need for using a particular system (Davis, Bagozzi and Warshaw 1989). Both of these factors are influenced by a number of other criteria and they are partly illustrated by Figure 3. Those factors printed in bold such as Navigation, Interactivity, Services are influenced by the system whereas the three italicized terms cannot be influenced by the system i.e. expectations, experience or can only be partly determined like speed.
TAM was supported by several studies (e.g. Lederer et al 2000) and concepts can be found in the literature extending the original TAM model and adding additional factors to the previously named perceived usefulness and ease of use, like perceived effectiveness and perceived success (Jung and Butler 2000), perceived accessibility and the customer’s attitude towards the Web (Jeong and Lambert 2001).

Wöber et al. (2002) introduced a framework distinguishing three categories of factors influencing perceived usefulness and perceived ease of use, i.e. personal factors (expectations, experience), system factors (such as product, speed, intelligence, services) and media factors (speed, operability). Personal and Media factors are uncontrollable to the system and the management whereas system factors represent controllable criteria.

Cheung, Chang and Lai (2000) found out that facilitating conditions and social factors are the two factors having a major influence on WWW usage. Facilitating conditions are especially relevant to the topic of recommendation systems since they describe the availability of necessary resources and supports given by the system used.

Another concept dealing with human-computer interaction but differentiating from the Technology Acceptance Model is the Concept of Flow (Novak, Hoffman and Yung 2000). Flow is described as a state of mind where the user is completely devoted to the use of a system and forgets everything else around him like time. Thus, the aim is to create a compelling online experience to facilitate flow. Figure 4 illustrates factors determining the cognitive state of flow.
These are some examples of the concepts relevant to computer-mediated environments in general and to recommendation system and therefore the DieToRecs project as well. Similar questions as the ones named above, arise when thinking of an evaluation of the first prototype. Which are the precise factors affecting perceived usefulness and perceived ease of use? The task success depends obviously on the type of recommendation system, its functions, the GUI (Graphical User Interface) and the navigation possibilities it has, the reliability the system dispose of or the interactivity quality it has. The number of cases, their spread and quality will have an influence on the behavior of the system as well.

2.3 Usability Testing and System Evaluation

The third theoretical area relevant to this project is usability testing and system evaluation. Numerous definitions for usability exist and a short overview will be given. According to ISO 9241-11 (1998) usability is “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction”. Lindgaard (1994) described usability as the ease of learning and using computer systems for experienced and inexperienced users.

Usability plays a central role when evaluating a system. According to Dillon (2003) the term “Usability Testing” in a general sense means all kind of UI inspection methods such as heuristic evaluations, expert review and cognitive walkthroughs. Usability testing in a more specific sense describes an evaluation procedure which comprises the performing of tasks by users.

Lindgaard (1994) distinguishes various data collection methods for usability testing exist being classified by the time of collection and if it is accomplished in the field or laboratory.
3. METHODOLOGIES

The evaluation procedure of the first prototype will be accomplished in April and May 2003 after the first prototype has been fully implemented. From the above named data collection methods, experiments further described in the following paragraphs will be employed to accomplish the evaluations. Two basic forms of data will be collected: objective and subjective measures. “Objective” data include the tracking of data using the system log functionality. Data about the user input, the information displayed, the interaction with the recommendation functions, system errors, performance and time measures will be collected. Table 1 illustrates this tracking of data by indicating some concrete examples for each of the areas named above.

<table>
<thead>
<tr>
<th>Table 1. Tracking of data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>User Input and Search:</strong></td>
</tr>
<tr>
<td>the number of queries</td>
</tr>
<tr>
<td><strong>Interaction with the Rec. Functions and System Errors:</strong></td>
</tr>
<tr>
<td>number of suggested relaxation operations</td>
</tr>
<tr>
<td>number of system errors</td>
</tr>
<tr>
<td><strong>Information displayed:</strong></td>
</tr>
<tr>
<td>number of results for each query</td>
</tr>
<tr>
<td><strong>Performance and Time Measures:</strong></td>
</tr>
<tr>
<td>Total task time</td>
</tr>
<tr>
<td>Number of items added to the travel plan</td>
</tr>
</tbody>
</table>
Both of the approaches cover some aspects of the broad concept of usability. According to Lindgaard (1994) computer system prototypes can be useful to conduct usability tests. There are several reasons why they should be conducted during the development of a system: the establishment of Web sites is expensive, thus they should be created the best way possible. Designers do not always know the needs and wants of users, users will reject systems not meeting their wants and finally, usability testing is a rather inexpensive way to test a site before its implementation (Boling 1995). The evaluation will consider some important preconditions for assessing adaptable and adaptive systems, i.e. real-world inquiries and user attitudes, e.g. the user’s satisfaction (Akoumianakis, Grammenos and Stephanidis 2001).

Experts have done a first evaluation of the already existing prototype. The method used was a cognitive walkthrough to give first hints which features could be changed or better implemented. The user evaluation will take place in April and May and will be done in three different countries: Austria, Italy and the U.S. There will be different variants of user testing, e.g. user evaluations of both approaches will be undertaken, and recommendation-oriented users are tested in the preference-oriented version and vice versa. Users will be assigned into different experimental groups and have to perform a training task with the system. Two variants of recommendation systems will be tested successively. Before they start with the training task, an introduction will be given by pre-trained staff and test persons are asked to write a short story about what they expect from the recommendation system. The assessment by users consists of a paper-and-pencil survey that will be filled out by the users themselves after they have experienced the system and have done a concrete task. Several surveys used in previous studies were taken into consideration when constructing the survey instrument: PSSUQ – The Post Study System Usability Questionnaire (Lewis 1995), SUS – The System Usability Scale (Brooke 1986), QUIS – The Questionnaire for User Interaction Satisfaction (http://www.cs.umd.edu/hcil/quis/) and finally, SUMI – The Software Usability Measurement Inventory (http://www.ucc.ie/hfrg/questionnaires/sumi/). However, the basis of the questionnaire is formed by PSSUQ, because it covers major aspects relevant to the DieToRecs system and it is enhanced by questions taken from the other approaches to make the survey more complete. The dimensions, e.g. design/layout, functionality, ease of use are hypothetical. A further research question could be to confirm or reject if this dimensions do form the influential factors of usability.

Table 1 shows the questionnaire concerning the system, user profile data such as age, gender, etc. are collected as well. Answers can be given on a 1-5 scale. The questions indicated in Table 2 are only general ones, there will be more specific questions concentrating especially on the recommendation functions and which will deal with both variants of the system: the idea pool and the preference-oriented version of the system.

<table>
<thead>
<tr>
<th>Table 2. Usability and User Satisfaction Questionnaire (adapted from PSSUQ)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Design / Layout</strong></td>
</tr>
<tr>
<td>I liked using the interface of the system.</td>
</tr>
</tbody>
</table>
| The organization of information on the systems screen was clear.
| The interface of this system was pleasant.                    |
| **Functionality**                                            |
| This system has all the functions and capabilities that I expect it to have. |
| The information retrieved by the system was effective in helping me to complete the tasks. |
| The list of item recommendations supported travel planning a lot. |
| I would have missed the function “complete travel recommendations” a lot. |
Ease of Use

It was simple to use this system.
It was easy to find the information I needed.
The information (such as online-help, on-screen messages, and other documentation) provided with this system was clear.
Overall, this system was easy to use.

Learnability

It was easy to learn to use the system.
There is too much information to read before I can use the system.
The information provided for the system was easy to understand.

Satisfaction

I felt comfortable using this system.
I enjoyed my travel planning session with this system.
Overall, I am satisfied with this system.

Outcome / Future Use

I was able to complete the tasks quickly using this system.
I was able to efficiently complete the tasks using this system.
I could not complete the tasks in the preset time frame.
I believe I could become productive quickly using this system.
The system was able to convince me about the goodness of the recommendations.
From my current experience with using the system, I think I’d use it regularly.

Errors / System Reliability

Whenever I made a mistake using the system, I could recover easily and quickly.
The system gave error messages that clearly told me how to fix problems.

4. EXPECTED RESULTS AND OUTLOOK

Since this is a work-in-progress and results are not yet available only expected results can be described by now. The outcomes of the assessments in April and May 2003 will be analyzed if there were factors found with which the users are unsatisfied and these features will be subject to change. This analysis should lead to improvements to be implemented in the second prototype. Furthermore, the second prototype will include three language versions, English, Italian and German and will provide data about Austria and Italy. Additional features such as user profiles, the possibility to store information and to get a complete travel plan, will be implemented\(^1\). This second version will be assessed as well, will be improved and will lead to the final recommendation system.

Obviously, the goal pursued with both of the evaluations is to achieve a high quality destination recommendation system. The assessments should lead to improvements in the GUI, the functionalities of the system, the ease of use. Were there some major problems detected in an area, e.g. the error probability or system reliability? If there is an obvious, often recurring problem, discovered, it has to be changed for the next prototype. Furthermore, the testing should reveal if some features are useful

\(^1\) DieToRecs Workplan
and will be therefore implemented in the final version or not. An example is the “idea pool”. How is this approach perceived by the (recommendation oriented) user? Is he/she interested in getting proposals without previously stating any preference or constraint? Another important question arising is that if the system is based on enough cases (in terms of quantity, quality and spread) to allow a useful similarity based recommendation. Furthermore, the results should reveal some facts which can only be detected by real interactions and which will lead to an optimal tourist recommendation system.

The goal behind the evaluations of DieToRecs and the major research question is to define which explanation model of the ones mentioned in chapter 2.1 about Human-Computer Interaction is most suitable for the case of recommendation systems. Does the measuring instrument have enough explanation power; is the research design strong enough are further directions for research.

Another question to be answered at the end of the evaluation is concerned with the assessments of recommendation systems itself. Which is an appropriate method to evaluate recommendation systems? The above described user testing for instance has the drawback that the two systems evaluated are presented one after another and it is very likely that users rate the second system better than the first one because of learning effects. Experts’ opinions are surely a helpful method to discover system weaknesses but on the other hand, this kind of assessment implies that a system is tested by highly skilled users which might not be the major target group for a system.

5. CONCLUSIONS

The search for travels via the Internet is a complex task. On the one hand, there is an abundance of information and on the other hand, the users do not always have the experience or time to process a large amount of information. Therefore, travel-advisory systems are a very convenient way for users to get selected travels or single items presented by the system according to his/her preferences. Present systems often have some shortcomings and DieToRecs is an attempt to overcome these problems and to provide the traveler with more useful results tailored specifically to the needs and wishes of the user. Constraints and preferences are considered, interactive query management is used, CBR, similarities to other users and personalized, already stored information are other features implemented in the system. The ultimate goal of this novel system is to enable a more convenient tourist destination selection by giving personalized travel recommendations. Thus, the development of an effective system aiding users in the information search and decision process by considering their individual preferences seems useful and represents the ultimate goal of DieToRecs. Moreover, the tourist recommendation system takes different cognitive styles into account enabling a more convenient human-computer interaction. The system operates with methods allowing recommendations being tailored to the preferences of the user or Case Based Reasoning to name but a few of the utilized approaches. The assessment of the prototypes of the system is a significant work to be done. Objective and subjective measures will be employed and combined to carry out the assessment. A survey will be carried out after the end of an experiment to have the user’s point of view and to assure that the system becomes usable and attractive. Additionally, experts’ opinions will be considered as well and it is self-evident that the system is constantly monitored by the members of the DieToRecs project team. Further research to be done will include the definition of criteria being essential for usability and the evaluation of a recommendation system.
REFERENCES


