Constraint-based Recommendation in Tourism: A Multi-perspective Case Study

Dietmar Jannach1, Markus Zanker2 and Matthias Fuchs3,4

1Department of Computer Science
TU Dortmund, 44221 Dortmund, Germany
E-Mail: dietmar.jannach@tu-dortmund.de

2Institute of Applied Informatics, University Klagenfurt,
Universitätsstraße 65-76, 9020 Klagenfurt, Austria
E-Mail: markus.zanker@uni-klu.ac.at
(Contact author)

3The European Tourism Research Institute, Mid-Sweden University,
83125 Östersund, Sweden
E-Mail: matthias.fuchs@etour.se

4eTourism Competence Centre Austria (ECCA),
6020 Innsbruck Austria
E-mail: matthias.fuchs@etourism-austria.at

Abstract: In many business-to-consumer (B2C) e-commerce scenarios, recommender systems (RS) have been shown to be valuable tools both for the online customer and the merchant. Such systems help customers find interesting items in large product assortments, increasing the chance of immediate online purchases and fostering long-term customer loyalty. However, standard technologies from classical RS application domains, such as books and movies, cannot be directly adopted in the tourism domain. This paper presents a case study of a constraint-based RS that was integrated into a travel advisory system for an Austrian spa resort. The study analyzes the system and its environment from three perspectives. Firstly, technological aspects of system development and maintenance are discussed; secondly, corresponding to the supplier’s view, the end user’s perspective is analyzed based on the findings of a study of the system’s usability and the perceived customer utility. Finally, the effectiveness of the system’s ability to positively affect user behavior is evaluated and discussed. The findings show that constraint-based RS not only help to positively influence tourist behavior, but that such systems can be built cost-effectively when using appropriate knowledge acquisition and maintenance tools.

Keywords: Recommender systems in e-Tourism, constraint-based recommendation, conversational recommender systems, web personalization, case study research.

1 This paper builds on and significantly extends the work presented in (Jannach et al., 2007) and (Zanker et al., 2008).
Introduction & Background

A recommender system (RS) can be characterized as a software module which is capable of generating a personalized list of interesting items for the visitors of an e-commerce site. For online visitors, the value of such a system is in its ability to offer personalized information and its information filtering functionality. In addition to this core functionality that supports in particular customers making decisions in domains complicated by large product assortments, RS are also supposed to increase the quality of the perceived shopping experience. Depending on the technology used, RS can for instance also be used to point users to unexpected but interesting items by introducing serendipity into the proposal generation process (Herlocker et al. 2004).

From the viewpoint of the provider of an e-commerce platform, the addition of an RS to a web shop can be advantageous for the following reasons. First, the RS can be seen as an additional online service to the customer that may help to gain a competitive advantage and to establish long-term relationships with the customers who might return to the online shop after appreciating the system's proposals. Aside from these long term effects, recent studies also show that RS can effectively influence the buying behavior of online customers in the short term, leading to significant increases in sales and market share (Shani et al., 2002; Zanker et al., 2006; Dias et al., 2008). Over the last fifteen years, RS technology has been successfully applied in various domains and recommenders have been built both for quality-and-taste products such as books or movies (Linden et al., 2003) as well as for complex and structured items such as consumer electronics (Jannach, 2004) or investment products (Felfernig et al., 2007a). Examples of recent studies on the effect of RS on the buying behavior of customers include (Shani et al., 2002; Zanker et al., 2006; Dias et al., 2008; Hegelich & Jannach, 2009). Over the same period, the Internet had an enormous impact on the travel and tourism industry, in particular with respect to the way how customers search for information about possible locations, how they locate accommodation, plan trips or how they may compare offerings from differing suppliers. Today, the Web is the primary information source in the pre-travel information-gathering process in Western countries (Werthner & Ricci, 2004). According to recent studies, e.g., by the European Travel Commission (ETC), customers are increasingly willing to purchase travel products online. This fact further contributes to the increased importance of online presence within the tourism domain. Consequently, over the last decade a large number of new online travel platforms have appeared which in turn has led to a highly competitive market place where it is hard to attract new customers and, more importantly, turn lookers into buyers. For a more detailed discussion of the role of web-based commerce in the tourism industry, see (Werthner & Ricci, 2004).

Given this highly competitive market structure, the integration of a value-adding RS, for instance into a booking platform, can be seen as a promising way to differentiate a site from its competitors and to increase market share and sales.

However, the most successful recommendation technologies used in practical settings, collaborative filtering and content-based filtering (Adomavicius & Tuzhilin, 2005), cannot be easily adopted to the travel and tourism domain for several reasons. Collaborative filtering (CF), as used for instance by Amazon.com for book recommendations, relies on the existence of a relatively large user community and the
idea that customers that liked the same items in the past will also have a similar taste in the future. In such systems, the prediction of whether a given customer will be interested in a certain item depends on his or her past rating and purchasing behavior and the taste of the community. When contemplating a recommender for travel arrangements, the problem arises that no detailed purchase histories or preference profiles are available for the majority of users. In addition, in particular for smaller tourism portals, e.g. for a certain geographical region, no large user community will exist, which however is a major prerequisite for the success of a CF-based RS.

For these reasons, knowledge-based and conversational approaches (Carenini et al., 2003; Felfernig & Burke, 2008) to product recommendation appear to be more promising in the tourism domain. In contrast to, e.g. collaborative filtering approaches that rely on past user behavior, knowledge-based approaches exploit domain expertise and means-end knowledge to determine a set of potentially suitable items for the customer. Thus, neither large user communities nor long purchase histories are required. In this regard, conversational recommendation means that end users should be given the opportunity to express their particular requirements with respect to the desired offerings (Thomson et al. 2004; Ricci et al. 2002). A typical example of a customer preference could be the price range for a hotel or particular equipment features in the hotel room. Overall, knowledge-based approaches have the advantage that the expert's knowledge is directly encoded in the RS and that neither a cold-start phase nor large user communities are required. As with all knowledge-based systems, however, the development of such a knowledge-intensive application requires some additional ramp-up effort in both knowledge acquisition and engineering, as well as in the development of an interactive user interface.

In this case study, we report on practical experience gained from the successful real world implementation of a knowledge-based, conversational tourism recommender system ("virtual advisor") developed for a premium spa resort in Austria. The application called VIBE has been online since 2005 and has been used by several thousand users since then.

The paper is structured according to the differing viewpoints of the various stakeholders involved with the design and operation of the application (i.e. the system developer's perspective, the end user’s view and finally, the perspective of the service provider and application owner - the spa resort). Accordingly, issues related to knowledge acquisition and representation as well as recommendation generation, aspects regarding the business value of the system, as well as empirically measured perceptions of end user utility will be addressed. The contribution of this real-world case study lies in its comprehensive discussion of the multiple perspectives that contribute to the successful implementation of a RS technology in the tourism domain.

In the next section, we describe the technological foundations and innovations of the application in the context of previous approaches in the field. The subsequent sections then discuss the perspectives of various stakeholders in more detail.
Constraint-based recommendation in tourism

The core problem in many recommendation scenarios in the tourism domain (e.g. destination selection, accommodation search or the selection of pre-packaged travel arrangements) is to find a set of items from a predefined catalog that match a set of given customer requirements. Table 1 lists fictitious holiday packages offered by a spa resort.

<table>
<thead>
<tr>
<th>Package Name</th>
<th>Min. duration</th>
<th>Included Activities</th>
<th>Room category</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sports weekend</td>
<td>3</td>
<td>Tennis facilities, swimming pool, golf court</td>
<td>3-star</td>
<td>299</td>
</tr>
<tr>
<td>Recreation week</td>
<td>6</td>
<td>Tennis facilities, health check, sauna</td>
<td>4-star</td>
<td>400</td>
</tr>
<tr>
<td>Luxury package</td>
<td>5</td>
<td>Wellness package, sauna, health check, massage</td>
<td>5-star</td>
<td>1099</td>
</tr>
</tbody>
</table>

Table 1 Example catalog of available items

Although all commercial booking platforms are fundamentally based on such an item catalog (or item database) most of them do not feature a personalized recommender system currently. The state-of-the-art in user support for the user's information retrieval and decision-making process implemented by today's commercial platforms is mainly characterized by static search forms and item lists that can be sorted according to different product features or popularity-based rankings based on customer ratings (e.g. for hotels). Sometimes, pointers to additional information materials are given and more recently, some sites host a user community in the form of discussion forums. While static search forms that are based on pre-defined search criteria might be of some value to customers who know exactly what they are looking for, this interaction style has several drawbacks, in particular (Jannach & Kreutler 2007):

- Product-centric item retrieval: static search forms are often organized in a product-centric way. For example, users may have to explicitly specify the desired item feature values such as the room category or a maximum price. Aside from the fact that a compromise between too few and too many search criteria on the search screen has to be found, a human travel advisor would tend to ask customer-centric questions, for instance as to why the customer is interested in a spa resort vacation, receiving answers focusing on health or beauty treatments or an interest in specific sports.

- No support for unsuccessful user queries and for item ranking: receiving no matching items for a search query is frustrating for the end user. An intelligent RS should, thus, be able to deal with such a situation and, for instance, provide a list of items that fulfill as many requirements as possible (Mirzadeh et al. 2004; McSherry 2005; Jannach 2006).

- One-style-fits-all user interfaces: web user groups are very heterogeneous. Users can differ with respect to technical skills, their background knowledge in the tourism domain and also with respect to their preferred way of retrieving information. Static
fill-out forms might, for instance, easily overwhelm a casual Internet user. Instead a flexible user interaction style that is capable of taking specific user characteristics into account might be more appropriate (Höpken et al. 2008).

Different content- and knowledge-based recommendation approaches can in principle be applied to the described travel arrangement recommendation problem (Ricci 2002). The VIBE virtual spa advisor system described in this paper uses a combination of existing and newly developed techniques and follows a constraint-based item retrieval approach. Before we discuss how the VIBE system differs from previous approaches, we formally characterize constraint-based recommendation; however refer the reader to (Felfernig & Burke, 2008) for a broader discussion.

The recommendation task as a constraint satisfaction problem

According to Tsang (1993), a classical constraint satisfaction problem (CSP) can be described by the tuple \((V,D,C)\) where:

- \(V\) is a set of variables
- \(D\) a set of finite domains for these variables, and
- \(C\) is a set of constraints that describes the combinations of values that the variables can simultaneously take (Tsang, 1993).

A solution to a CSP corresponds to the assignment of a value to each variable in \(V\) in a way that all constraints are satisfied. The recommendation problem can thus be expressed as a CSP as follows:

- \(V = V_C \cup V_P\), where \(V_C\) is the set of single-valued or multi-valued variables that capture user interests and \(V_P\) is a set of variables describing item features.
- \(D\) contains the domain definitions for variables in \(V\).
- \(C = C_P \cup C_I \cup C_F \cup C_O\), where \(C_P\) is a set of constraints for variables of \(V_P\) that describe the available items, \(C_I\) is a set of constraints capturing user requirements, and \(C_F\) is a set of filter constraints involving variables from \(V_P\) and \(V_C\) describing how customer requirements relate to product features. Finally, \(C_O\) is a set of other constraints that can be used, e.g. to express additional domain knowledge or to model inconsistent variable assignments for \(V_C\).

Motivating example. In our spa recommendation setting, the set of possible user inputs can be described using the variable set \(V_C = \{\text{max\_price}, \text{desired\_room\_category}, \text{desired\_activities}, \ldots\}\). The specific requirements \(C_I\) of a fictitious user could be, for instance, \(\{\text{max\_price} = 1000, \text{desired\_activities} = \text{Outdoo}, \text{desired\_room\_category} = \text{medium}\}\). Finally, the product features \(V_P\) are \(\{\text{package\_name}, \text{min\_duration}, \text{included\_activities}, \ldots\}\).

The item catalog \(C_P\) can be modeled as follows:

\[
C_P = \{ \\
\text{(package\_name = sports\_weekend} \land \text{min\_duration} = 3 \land \text{room\_category} = 3) \lor \\
\text{(package\_name = recreation\_week} \land \text{min\_duration} = 6 \land \text{room\_category} = 4) \lor \ldots \\
\} 
\]
Possible filter constraints in $C_F$ are:

- $\text{desired\_room\_category} = \text{medium} \Rightarrow \text{room\_category} > 2 \land \text{room\_category} < 5.$
- $\text{desired\_activities} = \text{Outdoor} \Rightarrow \text{Golf} \in \text{included\_activities}.$

In addition, a constraint could be added to $C_O$ expressing that some requirement combinations are not realistic, e.g.,

- $\text{desired\_room\_category} = \text{high} \land \text{max\_price} < 300$

For many recommendation problem settings, the constraint satisfaction problem can be simplified to finding a variable assignment for the variables in $V_P$ given a set of user inputs $C_I$. Every possible assignment to the variables in $V_P$ corresponds to a valid item recommendation. This problem solving task can be accomplished by any off-the-shelf finite-domain constraint solver.

Considering the above-mentioned limitations of today's mainly collaborative commercial systems, the presented constraint-based approach has the following advantages:

- It helps to overcome the problem of product centricity by allowing the user to express functional requirements over variables in $V_C$, relating them internally to (technical) item features.
- Item ranking functionality can be implemented by defining a corresponding optimization function and by using the typical branch-and-bound search functionality of modern constraint solvers.
- The concept of soft constraints, which is available in many constraint solvers, helps to resolve the problem of empty result sets for a given set of user requirements.
- While the user interface issue is left unresolved through a CSP-modeling approach, the framework permits possible situational user requirements to be modeled in different ways (e.g. to dynamically adapt the questions to the user based on her expertise or background knowledge). In addition, the constraints on allowed user requirement combinations in $C_O$ can be used to guide the end user through the requirements elicitation process.

**Comparison with other knowledge-based approaches**

The border between knowledge-based recommendation systems and classical expert or decision support systems cannot always be strictly defined, because decision support systems automatically providing expert advice to customers in online decision situations could in principle be seen as knowledge-based RS if they do consider personal characteristics and the situational context. Nevertheless, we focus on techniques that are subsumed under the term "knowledge-based approaches" in RS literature, namely utility-based ranking, critiquing and case-based reasoning.

In utility-based ranking, every catalog item is evaluated using different dimensions, which is typically accomplished by defining a function that assigns utility values to corresponding feature values (von Winterfeld & Edwards, 1986). In our example, a simple utility function could return higher utility values for higher room categories. In order to recommend an item, the sum of the utility values for all dimensions of all items is calculated and the item with the highest utility value is recommended. Since such recommendations are non-personalized, an extended model based on the Multi Attribute
Utility Theory – MAUT (von Winterfeld & Edwards, 1986) could be used, in which the user's specific interest in the given dimensions is taken into account.

The VIBE system described in this paper combines the constraint-based and MAUT-based approaches to provide cascading hybrid recommendations (Zanker et al., 2007). Firstly, a set of suitable items is determined using the given filter constraints; following this, the items are ranked according to their utility values. While the implementation of an optimization function as described above is in principle also possible in constraint-based approaches, the integration of the MAUT-based approach in a hybrid system has the particular advantage that the modeling process is simpler with utility functions.

Critiquing (Burke, 2002; McCarthy et al, 2006; Ricci & Nguyen, 2007; Reilly et al., 2005) is a recommendation approach in which the user is presented with an item recommendation on which he can give feedback about different product attributes. In the spa advisor scenario, the system could for instance propose the user a mid-range travel package along with the opportunity to give feedback in the form of "higher/lower room category" or "lower price". The domain-specific knowledge required in such approaches lies in the selection of the dimensions and features on which the user can give feedback. In addition the increment size for each dimension has to be determined. Recent improvements in critiquing technology include incremental, dynamic and compound critiquing, as described in e.g. (McCarthy et al, 2006; Reilly et al., 2005).

Critiquing functionality can be easily integrated into the constraint-based formalism by interpreting or translating every critique into a corresponding constraint on item features. In the VIBE system, however, no critiquing-based user interface was implemented. Instead, a forms-based and personalized question-answer dialog style was adopted as a richer way of eliciting the real user preferences (Jannach & Kreutler 2007).

A further possible way of recommending items based on domain-knowledge is based on case-based reasoning (CBR) technology, as reported in e.g. (Ricci et al., 2002; Ricci & DelMessier, 2004; McSherry, 2005). CBR is based on the idea of solving a new problem based on the solutions of similar cases of the past. The problem, therefore, consists somehow of retrieving a set of past cases that are sufficiently similar to the current case. The domain knowledge, therefore, mostly lies in the design of the similarity function. In many of the works mentioned above, however, a rather broad interpretation of CBR is used and a query-based (McSherry, 2005; Ricci & Werthner, 2002) retrieval method is typically used in combination with CBR methods.

Overall, the problem solving procedure of CBR and the constraint-based approach proposed herein are very different. Still, in principle, it would be possible to combine CBR and constraint-based recommendation by storing the past CSP solutions, later retrieving from the repository and adapting them in CBR style. In the VIBE system presented in more detail in the next section, however, no such functionality has been implemented.

Finally, note that – depending on particular requirements in the domain – special problem solving procedures might be required. A typical example is the problem of group recommendations or the problem of automated "bundling" of individual items into a single travel itinerary. For a discussion of such aspects, see (Ricci, 2002).
In the following sections, the VIBE case study will be discussed in detail and from multiple perspectives. Our primary goal is to focus on issues that arise in practical real-world recommendation scenarios in tourism.

The system developer’s perspective

Technical infrastructure – the ADVISOR SUITE system

The VIBE virtual advisor application is based on the ADVISOR SUITE framework, a now-commercialized software solution for the rapid development of conversational knowledge-based recommendation applications (Jannach, 2004; Felfernig et al., 2006, 2007a). Figure 1 presents an overview of the different components of the ADVISOR SUITE framework and the development process. The framework is based on a model-driven approach to web-application development in which the domain expert and a knowledge engineer specify the relevant domain knowledge with the help of user-oriented graphical knowledge editing tools. Subsequently, these models are translated into an executable web application as far as possible in an automated way. In order to make the system usable also for non-IT-experts, the system uses domain-specific and non-technical graphical notations. In addition, different views of the application (e.g. dialog modeling and constraint modeling) were introduced in order to reduce the complexity of the modeling task.

The acquired domain knowledge is stored in a central repository and, based on the definitions and the layout specification be provided by a Web designer, the system is capable of generating an executable Web application which, for instance, also includes the Web forms to interactively elicit user requirements: Examples of such elicitation forms are given in the sub-section on the end user's view. More details about the technical implementation of the framework can be found in (Felfernig et al., 2006).

Figure 1 Process and component overview, see also (Jannach & Kreutler; 2007).
Knowledge engineering and system development experiences

Experiences. While knowledge-based recommendation approaches offer advantages such as the ability to explain proposals to the user, they require the explicit formalization of the domain knowledge. In addition, the quality of the generated recommendation lists solely depends on the quality of the acquired knowledge. A structured acquisition and maintenance process and appropriate tool support is therefore crucial for the success of such a system (Felfernig et al. 2006).

With respect to the provided tools and modeling approach, experiences with the VIBE application were in general similar as in other applications developed based on ADVISOR SUITE's rapid application development approach (Jannach & Kreutler, 2007). The system's tools and notations for model development for instance could be used for developing the system in an incremental and evolutionary fashion. While the domain experts did not manipulate the models by themselves in this application and the expert was always assisted by a knowledge engineer, the system's capability to automatically generate working web applications from the models was particularly helpful to validate the models with the domain expert. In addition, we also found that the domain expert's involvement and commitment to the project increases when suggestions are immediately reflected in the developed software.

With respect to the size of the knowledge base, the evaluation showed that a few dozen constraints were sufficient for capturing the main aspects of the recommendation and personalization logic. The size of the knowledge base, therefore, remained at a manageable level. In addition, the modular design of the knowledge base and the corresponding separation of different views further helped the domain experts to keep track of the current process state.

Overall, our experiences were in line with the experiences obtained from previous applications, i.e. system development and maintenance costs can be significantly reduced compared with traditional web application development approaches, in particular because changes in the knowledge base are directly reflected in the user interface of the ADVISOR SUITE system (Jannach & Kreutler, 2007).

Limitations and challenges. The tourism domain, however, also has particularities that are not found in other recommendation domains. An example is the fact that tourism offers are often available only in certain periods of time. Such functionality was not part of the ADVISOR SUITE system and required extra treatment, in particular because the time constraints (i.e., user-specified start and end times) in most cases cannot be treated as hard constraints. As time constraints cannot be encoded as soft constraints, a domain-specific constraint type had to be implemented. This procedure allowed the RS to take into account touristic offers that did not exactly but only partially match the user's time preference for the trip. Fortunately, the extensible design of the ADVISOR SUITE system allowed us to implement such a domain-specific functionality without changing the core item retrieval and constraint relaxation algorithm. The formal integration of the concept of flexible time constraints in the CSP-based approach to recommendation remains part of our future work.

Note that with respect to the relaxation technique, the method implemented in ADVISOR SUITE provides similar functionality to the Intelligent Travel System (ITS) described in
(Ricci & Werthner, 2002, 2004). In addition, ITS also supports "query tightening" as part of its interactive query management component. In situations, in which too many items match the customer requirements, the system generates a proposal for the user, suggesting which item features should be further constrained to narrow down the set of relevant items. While such functionality can be helpful in tourism recommender situations with large catalogs (e.g. hotel recommenders), in our setting this was not required because of the limited catalog size.

Other particular aspects of the tourism domain for which more research is clearly required include the problems of group recommendations (traveling is most often a group decision and group members may have different preferences) and the problem of automatic bundling or packaging of items (Zanker et al. 2009). While the first issue has not been addressed in the VIBE application, the second requirement was solved in this project. Currently, we are investigating how existing product configuration technology, see for instance (Mittal & Frayman, 1989; Mailharro, 1998) can be embedded in the constraint-based recommendation scheme.

Finally, we anticipate further research in knowledge-based travel recommendation in the area of self-adapting or learning knowledge-supported systems that use a combination of explicitly-provided domain knowledge and learning functionality. Current tourism recommender systems are mostly knowledge-based. One of the few exceptions is the system presented in (Wallace, 2004), that, in contrast to standard collaborative filtering methods, not only takes the purchase history of the individual customer into account but also tries to learn general usage patterns from the complete transaction log. While this approach has been shown to be beneficial when usage data is sparse, it cannot profit from explicit domain knowledge. Therefore, we aim to develop approaches that can benefit from the advantages of both worlds. Details of a first method which is capable of automatically extracting constraints from historic interaction data can be found in (Zanker, 2008).
The end user’s perspective

Web application user interface

End user access to the VIBE virtual advisor is provided via the multi-lingual web portal of the spa resort. Upon receiving a user request, the virtual advisor is started in a wizard-style window, in which the user is guided through an interactive sales and advisory dialog. The dialog is conducted using the female avatar VIBE (signifying the name of an ancient goddess of springs), a life-like avatar which helps to increase the system's persuasiveness by means of "personification" (Zanker et al., 2006).

Figure 2 Welcome screen of the virtual advisor

Figure 2 shows the welcome screen and first question asked by the advisor. The general interaction scheme of the application can be summarized as follows. First, the user's explicit requirements and preferences are elicited in a system-guided and form-based dialog. During the dialog, the system constantly analyzes the user inputs and updates the internal user model. The selection of the next dialog action – such as asking the next question or interrupting the dialog for an additional hint – is determined based on the current user model and the dialog personalization knowledge modeled during the knowledge acquisition phase. While a discussion of the dialog modeling and personalization component is beyond the scope of this paper, details can for instance be found in (Jannach & Kreutler, 2007). The wizard-style interaction with “next” and “back” navigation options allows users to reconsider and revise their preferences, in particular when the system detects inconsistent user requirements.

Once sufficient requirements and preferences have been elicited, the advisor displays the best matching proposals (Figure 3) by filtering those items that fulfill the user's constraints. With respect to navigation opportunities, the user can navigate back and revise the requirements or ask for an explanation or more details about the proposed
items. The provided explanations are based on the set of filter constraints $C_F$, which were annotated with natural language explanatory statements in the modeling phase. If not all constraints can be satisfied, the system also displays the set of constraints which have been automatically relaxed.

In addition to the individual recommendations, the user is also presented with a set of optional add-on packages as shown in Figure 3. The selection of these additional items is based on a similar constraint-based selection technique. As mentioned, this "bundling" functionality was implemented as a domain-specific add-on to the ADVISOR SUITE system.

Overall, the end user experience of the VIBE virtual advisor significantly differs from the searchable product catalogs which can be found on most of today's tourism platforms. In contrast to these "passive" web applications, the VIBE advisor actively guides users through a personalized preference elicitation dialog which focuses on customer preferences rather than on product features.

Utility and usability analysis

The more complex a purchase decision is the more a website can facilitate exploratory browsing. Similarly, a RS is expected to accomplish this purpose without producing friction or annoyances that provoke users to quit the application. Moreover, an interactive travel advisor is expected to support users (i.e. potential tourists) in navigating through a large space of electronically offered travel service packages (Zanker et al. 2008, p. 27). Thus, perceived utility and the application’s usability represent appropriate indicators to approximate efficiency in retrieving relevant tourist information to effectively support travel decision making and booking processes (Felfernig et al., 2007b).
Below a click-stream sequence analysis is presented in order to detect eventual frictions that provoke users to quit the virtual travel advisor VIBE. Furthermore, online interaction paths as well as the relative share of traffic at each node are visualized (Senecal et al., 2005). The findings are based on web-usage data logged during the second half of 2006. A total of 712 distinct user sessions are analyzed. As can be shown, after question six users can switch either to the final page comprising individual recommendations or may further communicate preferences to the system (Zanker et al. 2008, p. 28). More importantly, however, it became evident that despite 58.6% of all users reached the recommendation page, the majority left the system already within the first three dialogue steps (Fig. 4). Thus, there is a need to assess customer’s perceived usability and utility (Yaijng et al., 2007).

The goal of usability testing is to identify usability deficiencies existing in computer-based applications (Nielsen, 1993; Rubin, 1994). A first predominantly qualitative test took place in Innsbruck at 20th of July 2007. With the help of a test group comprising eight individuals (i.e. 4 male and 4 female participants between 20 and 63 years old) positive and negative experiences when using VIBE were gathered. In addition, the first navigation stage was recorded by a Tobii eye tracker (Duchowski, 2003) and analyzed with the Eye-Gaze software. Eye tracking provides a practical add-on to usability testing as it clearly shows which parts of a specific user interface are perceived and correctly recognized. Interestingly, any of the eight testers clicked on VIBE without external help. More precisely, only two subjects see VIBE, but they do not understand the purpose of it. Six people declare they did not recognize VIBE at all. However, a look on the heat maps tells the opposite. Each subject looks at VIBE at least once. For instance, participant I states that she didn’t see VIBE, but figure 5 (left) proves that she looked at VIBE. Figure 5 (right) shows the hot spot of participant II. He is highly interested in the portal and looks at every detail. However, during the unstructured post-test interview he explains that he saw VIBE, but couldn’t understand its purpose. Moreover, since he searched for vacations for his family he looked for arrangements for his children and VIBE did not appeal suitable for this matter.
In the course of unstructured post test interviews the following usability and utility aspects were mentioned. All participants positively assess the usability of the system and can easily handle it and don’t have any navigation difficulties. Especially the women metaphor and the linear step-by-step procedure effectively support the recommendation dialog in the eyes of the testers. Also the multiple choice questions are described as being easy and quick to answer. Similarly, also the personal information to provide to the RS is not criticized. The individualized recommendation is the mostly pronounced positive utility aspect. All participants state that they appreciate that they did not have to search through the whole website, but instead had a tool to browse for ideal holiday packages that considers their preferences best. Finally, the majority (i.e. 7 out of 8 testers) would recommend VIBE. Study participants, however, also detected various problem fields. The most negative aspect is the fact that 6 out of 8 participants did not perceive VIBE on the starting website of the thermal spa resort. Six testers explicitly state the unfavorable position on the website. Furthermore, VIBE’s comments via the woman metaphor are not perceived as interesting and nearly anybody reads it. All testers state that the prices of the various tourism services did not clearly appear (e.g. price per person or price per room). Also the result list is rather difficult to read and there were displayed too many packages at once. Particularly, they criticize that one has to search for detail information on his own without direct hyperlinks to the website of individually recommended tourism suppliers. Finally, most testers don’t know what to do after the recommendation is completed and the majority confuses the term “online request” with reservation, booking or possibility to send a reservation e-mail. Interestingly, every study participant stated that VIBE should also support the reservation and booking process.

Since the System Usability Scale (SUS) proves to be a valuable evaluation tool being robust and reliable with relatively small samples (Brooke, 1996), in the next analytical step SUS was used as a quantitative measure for usability testing. SUS is a ten-item Likert type scale allowing respondents to subjectively assess the usability of an electronic system. Typical SUS items are “I would like to use this system frequently”, “The system is unnecessarily complex”, “The system is easy to use”, “I found the functions in this system were well integrated” and “Most people would learn to use this system very quickly”, etc. The usability test was conducted in the course of an Email survey during the first two weeks of August 2007. In total 75 emails were sent to randomly selected adults in Austria and Germany with the request to use VIBE and to return a completed questionnaire. The task was to plan a vacation for oneself and a partner or the family at
the thermal spa resort using VIBE. After one follow up email after the first week, a total of 55 questionnaires returned filled out by 26 female and 29 male persons aged between 19 and 67 years (i.e. average age is 38 years). Two-thirds (65.4%) identified themselves as advanced computer users while one-third (34.6 %) are intermediate or beginning users. SUS yields a single composite score for overall usability of the test system (Brooke, 1996) ranging from 0 (i.e. least usable) to 100 (i.e. most usable). Thus, to obtain a SUS score value individual contributions (i.e. ranging from 0 to 5) are summed up and multiplied by 2. For the interpretation of SUS scores, threshold values are recommended by literature (Brooke, 1996). Score values between 80 and 100: users like the system; scores between 60 and 79: users accept the system and scores below 60: users refuse the system. The empirical result revealed 81.5 SUS points, thus, the participating subjects seem to like VIBE.

The behavioral intention to use an information system is determined by an individual’s attitude towards that system in terms of beliefs about its usefulness (i.e. utility) and usability (i.e. ease of use) (Davies, 1989; Rubin, 1994; Davies & Venkatesh, 2004). However, in case of RS there is hardly any research concept and methodological approach that investigates user acceptance (Swearingen & Sinha, 2001; Zins & Bauernfeind, 2006). Thus, the below presented acceptance test combines a usability testing model with a utility measurement. For this purpose, the model integrates test items from Brooke’s (1996) SUS Model and Davis & Venkatesh’s (2004) technology acceptance model. Data is gathered in the course of the online survey previously reported and revealed the following results: The item “VIBE is useful to support efficient information search” shows an average value of 2.69 on a 5-point Likert scale (i.e. 1 = I fully agree, 5 = I don’t agree at all). The overall measure for perceived usefulness, however, shows a mean of 3.44. This result clearly shows that relevant utility dimensions (e.g. productivity increase for the whole travel decision making and booking process, improved quality of search findings, time reduction, etc.) are not yet fully satisfied by the virtual travel advisor VIBE. Interestingly enough, although system’s usability (i.e. SUS) was assessed more favorably than overall utility, a multiple regression revealed that only the latter item significantly determines the intention to use the travel advisory system ($R^2 = 32\%$). Thus, the rather moderate test results for the perceived usefulness clearly pinpoint important improvement potentials to remove current barriers to use the travel advisory system (Davies & Venkatesh, 2004).

To sum up, VIBE reached fairly good usability results and the RS is accepted and liked by the test users which would use the application frequently and recommend it to other users. However, one of the biggest usage problems is that most subjects did not perceive VIBE on the web site of the thermal spa resort. Thus, VIBE should be better positioned on the online portal. Although in the eyes of the testers, VIBE increases the efficiency of information search, the utility analysis revealed only moderate results. Moreover, perceived usefulness proved to be the most important barrier to use the virtual travel advisor. Thus, in order to increase present usage rates with that system, the following suggestions to enhance utility dimensions are deduced from the study results: The virtual advisory system should continue to support availability checking and online booking processes, respectively. As people seem not to be pleased with the arrangement of VIBE recommendations, the latter should be reduced by number (i.e. max. top 3 packages) and classified by category. Finally, the “detail” link should be improved by including useful information, such as pictures, price information, main user preferences on which the
recommendation is based on as well as hyperlinks to the recommended tourism suppliers’ websites and offered travel arrangements.

The service provider's perspective

In the following section we will discuss the value of a knowledge-based recommender system from the perspective of the application owner and operator offering the recommendation service to its online clients. In this scenario the owner of the application and the website itself is also the producer of the tourism services that are recommended by the system. Therefore, the business goal of the application provider lies in harnessing the power of the Web in order to reach and communicate directly to customers. According to (Rabanser & Ricci, 2005) the application provider follows the “Manufacturer Model” and the recommendation system helps the spa resort (1) to increasing the quality of the online presence, (2) to better understanding the needs of customers and (3) to improve the efficiency of the website. The impact of a recommender system on the service quality perceived by users has already been explored in the previous section discussing the customer’s perspective.

An increased understanding of the needs of users is one additional positive effect derived from operating a recommender system and in particular a conversational one like in this case study. When users interact with the system they provide valuable feedback that can be exploited to improve prediction mechanisms as discussed in (Zanker & Jessenitschnig, 2009) or to improve the understanding of users’ needs and to gain additional marketing knowledge. The latter can be placed under the concept of marketing intelligence as proposed by (Büchner & Mulvenna, 1998). In the context of log data from an interactive advisory system the goal is to identify coherent customer segments and outstanding usage patterns that can lead to new service developments. (Zanker et al., 2008) proposed correspondence analysis to visualize relationships between non-metric attributes and to identify groups of corresponding characteristics.

Accordingly, the third aspect targets the effectiveness of an e-commerce site that is commonly measured using conversion metrics such as the lookers-to-bookers ratio or the

Figure 6 (Quasi-)experimental research design

Accordingly, the third aspect targets the effectiveness of an e-commerce site that is commonly measured using conversion metrics such as the lookers-to-bookers ratio or the
share of requests for proposals among total online visits. In order to evaluate the conversion criteria on real users we decided to pursue a quasi-experimental research design comparable to the case study of (Delgado & Davidson, 2002). They evaluated SkiMatcher, Ski-europe.com’s conversational recommendation system for ski resorts, over a period of 4 months in 2001.

In Figure 6 we depict the quasi-experimental research design where historic online visits are assigned to the treatment, i.e. interaction with the recommendation system, by self-selection. Admittedly, the assignment of users to the use of the recommendation system was not random. However, as the usability study in the previous section stated, most users were unaware of the availability of a RS simply because they did not recognize its entry button. As a consequence only around 2% of all users went through the interactive conversation, expressed their wishes and requirements and received recommendations. Thus, the mental processes that influence the cognition of the website’s functionality such as its RS are partly responsible for the assignment of participants to treatments. However, these processes cannot be measured and controlled in a field study and thus constitute natural selection.

<table>
<thead>
<tr>
<th>Sessions</th>
<th>Month1</th>
<th>Month2</th>
<th>Month3</th>
<th>Month4</th>
<th>Month5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>6525</td>
<td>7841</td>
<td>9971</td>
<td>9599</td>
<td>8175</td>
</tr>
<tr>
<td>RS Users</td>
<td>110</td>
<td>143</td>
<td>207</td>
<td>196</td>
<td>163</td>
</tr>
<tr>
<td>Non RS Users</td>
<td>6415</td>
<td>7698</td>
<td>9664</td>
<td>9403</td>
<td>8012</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RFP Generated</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>123</td>
<td>247</td>
<td>427</td>
<td>381</td>
<td>240</td>
</tr>
<tr>
<td>RS Users</td>
<td>8</td>
<td>11</td>
<td>18</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Non RS Users</td>
<td>115</td>
<td>236</td>
<td>409</td>
<td>366</td>
<td>225</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conversion</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>1.88%</td>
<td>3.15%</td>
<td>4.28%</td>
<td>3.97%</td>
<td>2.94%</td>
</tr>
<tr>
<td>RS Users</td>
<td>7.27%</td>
<td>7.69%</td>
<td>8.7%</td>
<td>7.65%</td>
<td>9.2%</td>
</tr>
<tr>
<td>Non RS Users</td>
<td>1.79%</td>
<td>3.07%</td>
<td>4.23%</td>
<td>3.89%</td>
<td>2.81%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Increase in Conversion</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RS Users/Non RS Users</td>
<td>406%</td>
<td>250%</td>
<td>206%</td>
<td>197%</td>
<td>327%</td>
</tr>
</tbody>
</table>

**significant p < 1%, *significant p < 5%**

<table>
<thead>
<tr>
<th>Yates $\chi^2$(1)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yates $p$</td>
<td>&lt;0.001</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Fisher 2-tail: p-value</td>
<td>&lt;0.002</td>
<td>&lt;0.005</td>
<td>&lt;0.005</td>
<td>&lt;0.02</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cramer’s V</td>
<td>0.0475</td>
<td>0.0327</td>
<td>0.0295</td>
<td>0.0253</td>
<td>0.0503</td>
</tr>
</tbody>
</table>

Table 2 Summary of results over 5 months, year 2006

In a preprocessing step historic user sessions (N = 42,111) were extracted from web-log data following the procedures for data preparation as discussed in (Mobasher, 2007). Weblog entries that derived from search bots as well as those entries that could not be unambiguously assigned to a session due to unresolved DNS names of accessing hosts or the use of proxy-servers were removed. The total number of sessions, the use of the RS and the conversion metric are reported in Table 2 for 5 monthly periods. In our context conversion means that the user submitted a request-for-quote (RFQ). Thus, the conversion metric reports the relative share of users submitting a RFQ compared to the total.
The null hypothesis $H_0$ denotes that the use of the recommender system and the submission of an RFQ are conditionally independent for each month and it can be rejected as a consequence with a p-value below 1% based on $\chi^2$ statistics including Yates’ correction for continuity. In addition we computed Fisher’s exact test for small numbers that results in a slightly lower level of significance and a p-value below 2% for the fourth month. Summarizing, the contingency analysis indicates that users of the RS are significantly more likely to convert than non-users. In order to evaluate the strength of this relationship between the two nominally scaled variables we computed Cramer’s $V$ which is the same as the Phi-coefficient for 2x2 dimensional tables. However, as the results in Table 2 indicate the strength of this relationship is rather low.

<table>
<thead>
<tr>
<th></th>
<th>Month1</th>
<th>Month2</th>
<th>Month3</th>
<th>Month4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sessions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>10714</td>
<td>15560</td>
<td>18317</td>
<td>24416</td>
</tr>
<tr>
<td>RS Users</td>
<td>1027</td>
<td>1673</td>
<td>1878</td>
<td>2558</td>
</tr>
<tr>
<td>Non RS Users</td>
<td>9687</td>
<td>13887</td>
<td>16439</td>
<td>21858</td>
</tr>
<tr>
<td><strong>RFP Generated</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>272</td>
<td>506</td>
<td>445</td>
<td>641</td>
</tr>
<tr>
<td>RS Users</td>
<td>75</td>
<td>143</td>
<td>161</td>
<td>229</td>
</tr>
<tr>
<td>Non RS Users</td>
<td>197</td>
<td>363</td>
<td>284</td>
<td>412</td>
</tr>
<tr>
<td><strong>Conversion</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2.54%</td>
<td>3.25%</td>
<td>2.43%</td>
<td>2.63%</td>
</tr>
<tr>
<td>RS Users</td>
<td>7.30%</td>
<td>8.55%</td>
<td>8.57%</td>
<td>8.95%</td>
</tr>
<tr>
<td>Non RS Users</td>
<td>2.03%</td>
<td>2.61%</td>
<td>1.73%</td>
<td>1.88%</td>
</tr>
<tr>
<td><strong>Increase in Conversion</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RS Users/Non RS Users</td>
<td>359%</td>
<td>327%</td>
<td>496%</td>
<td>475%</td>
</tr>
</tbody>
</table>

"*" significant with alpha < 1%

Yates $\chi^2$ (1)  
Yates p  
Fisher 2-tail: p-value  
Cramer’s V

Table 3 Summary of Delgado & Davidson (2002)’s results over 4 months, year 2001

In the next step we compare our findings to the results obtained by (Delgado & Davidson, 2002) as shown in Table 3. Note that we computed the missing statistical significance tests and relationship strengths based on their published numbers. Interestingly enough, our study results replicate their historic results. Average conversion rates of RS users as well as the relative increase in conversion compared to non-users were approximately the same. However, one major difference between the two cases lies in the usage rate of the recommender system among website visitors. While the SkiMatcher RS was accessed by approximately 10% of all users, VIBE was only used by 2% of total website users for the reasons discussed previously. Thus, the RS adoption rate was 5 times lower in the spa resort scenario, which also led to a lower strength of the relationship between RS usage and Conversion compared to the results from the SkiMatcher case given in Table 3.

However, questions of causality such as “does the use of a RS persuade users to convert?” cannot be fully answered by this study which constitutes a limitation. Nevertheless, this analysis of more than 40K historic user sessions showed that the use of a RS and the submission of an RFQ are significantly correlated over several months and
also replicated earlier results of (Delgado & Davidson, 2002) on a different website also featuring a knowledge-based conversational recommender system.

Conclusions

This paper presented a case study of using knowledge-based recommendation technology in the tourism domain. In the context of the VIBE virtual advisor, which was developed to help end users find an appropriate package at a thermal spa resort, the paper analyzed the recommendation problem setting and the implemented solution from the perspectives of multiple stakeholders.

With respect to recommendation technology, the usage of a constraint-based knowledge representation and reasoning framework was proposed to address the shortcomings of classical algorithms such as collaborative filtering and content-based filtering, which cannot be applied in the given problem setting where detailed customer purchase and preference histories are unavailable. From the perspective of the system developer, the availability of an adequate modeling environment was shown to be particularly beneficial in overcoming the knowledge engineering bottleneck which is typical in such knowledge-intensive applications.

From the end user’s perspective, VIBE reached satisfactory results based on a qualitative and a quantitative usability/utility testing approach (Rubin, 1994; Brooke, 1996). Although the RS is liked by all test users which would also recommend it to other users, various usability and utility problems were identified. Suggestions for system improvement are (1) the repositioning of VIBE on the online portal, (2) extending VIBE in order to support also online reservation and booking, (3) reducing the number of adequate recommendations (i.e. max. top 3 packages) and, finally (4) adding useful information in the detail link (e.g. pictures, price information, user preferences on which the recommendation is based on and hyperlinks to recommended tourism suppliers).

Finally for the application provider and operator (i.e. the spa resort) a quasi-experimental analysis that reconstructed historic user interactions from real-world log data over a period of several months showed that users of the conversational RS are more likely to issue a request-for-quote. Quite interestingly, this study replicates the earlier results of (Delgado & Davidson, 2002) who analyzed the conversion rate of users of the SkiMatcher application fielded on ski-europe.com.

Overall, we see our work as a further step toward gaining a better understanding of the different aspects of knowledge-based RS in the tourism domain for which few publicly available studies exist.

Acknowledgement: The authors wish to thank Nina Müller and Martina Fritsch for their useful support in data gathering and analysis.
References


