

Evaluating Recommender Systems in Tourism – A Case Study from Austria

Markus Zanker^a, Matthias Fuchs^b, Wolfram Höpken^b, Mario Tuta^b and
Nina Müller^b

^a Institute for Intelligent Systems and Business Informatics
University Klagenfurt, Austria
markus@ifit.uni-klu.ac.at

^b e-tourism competence center Austria (ECCA)
Innsbruck, Austria
{firstname.lastname}@etourism-austria.at

Abstract

Recommender systems (RS) are employed to personalize user interaction with (e.g. tourism) web-sites, supporting both navigation through large service assortments and the configuration of individual service packages. Depending on the interaction strategy, RS are either utilized to elicit users' tastes and preferences or to stimulate desire for different offerings. In addition, as a potentially rich source of digital traces, RS also act as a repository for marketing intelligence. Web-usage mining is an accepted approach to analyse web-usage behaviour based on information traces left by the web-user (Mobasher, 2007). This paper proposes an empirically tested approach which combines typical web-log data with user feedback gathered by an interactive travel advisory system developed for an Austrian spa-resort. The proposed approach focuses on evaluating the RS with respect to efficiency, effectiveness and actionable marketing intelligence.

Keywords: Recommender system, interactive advisory system, web-usage mining

1 Introduction

Web-usage mining aims to exploit vast amounts of transaction data in order to gain valuable marketing intelligence (Cooley et al., 1999). Today, a variety of standard tools exist (e.g. *123logalyzer*) that allow webmasters to effectively analyze user-driven traffic on their platforms (<http://www.123logalyzer.com>). These tools generate standardised reports comprising, for instance, the amount of unique visits, the geographic location of accessing clients or referring sites (Büchner & Mulvenna, 1998). No doubt, this information is highly relevant for determining bandwidth and the maximal load on the web-server; for deciding to offer, for instance, additional

multi-lingual content; or for improving search-engine marketing. However, it does not provide advanced insights into web users' behavioural patterns. Both Büchner and Mulvenna (1998) and Cooley et al. (1999) described methodological approaches for combining different sources of Internet data (e.g. web-server logs and marketing data) to discover actionable marketing intelligence. Similarly, the aim of the paper is to exploit additional information like the explicit feedback obtained from the users of an interactive travel advisor. Combining this data with data obtained from web-server logs facilitates the identification of user segments based on similar preferences or behavioural patterns.

The paper starts with a brief discussion on studies in the application field of both recommender systems in tourism and the corresponding web-usage data mining approaches. Motivated by the main success factors in the context of RS applications, namely efficiency, effectiveness and marketing intelligence, Section three proposes an evaluation approach which supports this scope. For instance, by analysing users' *click-stream sequences*, the interactive dialogues can gradually be improved, and thus be made more efficient (Senecal et al., 2005). Furthermore, the effectiveness of a RS application can be evaluated by *cross-tabling* non-metric user characteristics and computing corresponding contingency parameters (Giudici & Castelo, 2001). Finally, additional insight into the usage behaviour can be gained by visualizing the relationships between the various user-characteristics based on *correspondence analysis* (Hair et al., 2006). In Section four selected results are presented. Finally, the conclusion summarizes the findings and briefly outlines future research agendas.

2 Background

Due to the informational nature of travel and tourism services, the industry is the leading application field for electronic commerce (Werthner, 2002). Intelligent applications, like recommender systems, enable users to become their own travel agents. Ricci (2002) provides several examples where a matching engine derives recommendations according to the user input and, in addition, gives an excellent introduction into the field. *SkiMatcher* (Delgado & Davidson, 2002), for instance, is a well-known example that incorporates multiple recommendation paradigms, such as constraint-based, content-based and collaborative information filtering techniques. Following Burke's classification of recommendation techniques, constraint-based recommendation is a variant of knowledge-based RS exploiting explicit domain knowledge (Burke, 2002). Furthermore, content-based recommender systems reason on the similarity between product items in order to propose additional items to a user. In contrast, collaborative filtering techniques recommend items based on user similarities. Finally, case-based recommendation applications utilize techniques from all the three paradigms (Ricci & Werthner, 2002), first by asking the user to state

his/her preferences and then by improving over time and learning from past interactions (i.e. their case base).

The interactive travel assistant that acted as a case for our study can be assigned to the constraint-based recommender group (Jannach et al. 2007). Its implementation is based on the knowledge-based ADVISOR SUITE framework presented in (Felfernig et al., 2006). More precisely, it uses a single interaction style, i.e. a forms-based dialogue to ask the user for tourism motives and preferences, like relaxation, health improvement or different sports activities. Figure 1 depicts this question/answer style of interaction. The result from this interactive advisory process is a personalized list of recommended travel packages.

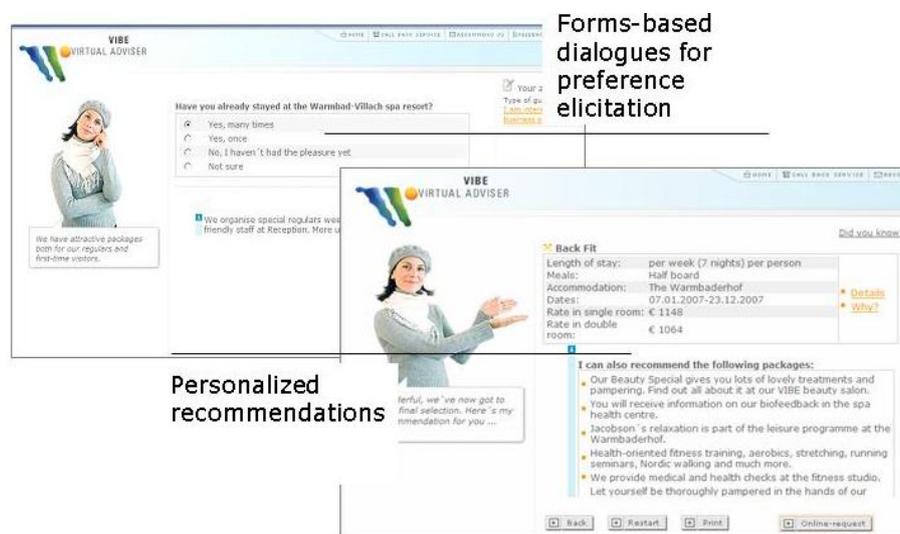


Figure 1 Screenshots of the interactive advisory system

With respect to the evaluation of recommender systems Herlocker et al. (2004) provided an encompassing overview. Furthermore, Pan and Fesenmaier (2003) investigated users' search behaviour from the viewpoint of their mental models and their semantic understanding. Similarly, Zins et al. (2004) conducted usability tests based on questionnaires to compare user satisfaction after using different destination recommender systems. In contrast, the proposed approach uses automatically collected data to derive insights into the system's performance in terms of efficiency, effectiveness and marketing intelligence, respectively. The proposed evaluation approach refers particularly to Mobasher's (2007) state-of-the-art paper on web-usage

mining techniques for personalization. It provides an overview on the necessary data collection and preprocessing tasks and describes the techniques for predictive user modelling.

3 Proposed Web-Usage Mining Approach

Supported by related work, data pre-processing is a highly necessary and important precondition when conducting web-usage mining (Cooley et al. 1997; Büchner & Mulvenna, 1998). Figure 2 sketches the proposed knowledge discovery process based on the WEBMINER architecture described by Cooley et al. (1999).

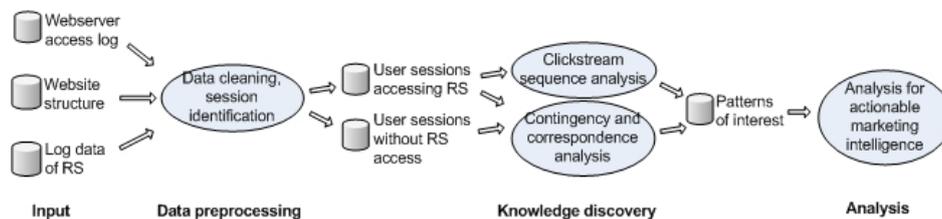


Figure 2 Knowledge discovery process (source: adapted from Cooley et al., 1999)

The knowledge discovery process starts with the integration of several data sources. Subsequently, in a data cleaning step, user sessions from the raw access log entries are firstly identified and all sessions deriving from web robots are removed. Secondly, if the users invoked the interactive travel assistant, their explicit answers from the logging facility of the advisor component are extracted. Thus, two data sets resulted from this pre-processing step: user sessions that accessed the recommender system and those that did not.

As already mentioned, the proposed evaluation approach focuses on rating the RS with respect to the dimensions *efficiency*, *effectiveness* and *marketing intelligence*. Herlocker et al. (2004) extensively discuss evaluation criteria for RS. They note that most evaluations focus on accuracy. However, in commercial environments goals like conversion rate, ease-of-use or creation of strategic knowledge become more important. Put differently, a RS is expected to accomplish its purpose without producing friction or annoyances that provoke the user to quit the application. Moreover, an interactive travel advisor is expected to support users in navigating through a large space of electronically offered travel service packages. Therefore, next to customer satisfaction the application's ease of use or more general its usability represents an appropriate indicator to approximate efficiency. Thus, to cope with this aim, a *click-stream sequence analysis* (Senecal et al., 2005) is employed. With respect to the effectiveness of a RS the achievement of its core objectives is of primary

interest. These are both click-through and conversion rates, respectively, as well as the capability to retain users on the site (Fesenmaier et al., 2003; Zins et al. 2004). Accordingly, in the advisory scenario analyzed below, a user visit is successfully completed by an online reservation request. Cross-tables and corresponding *contingency analyses* are thus proposed as appropriate means of determining significant relationships between non-metric variables, such as user-group characteristics and conversion rate (Pallis et al., 2005).

Finally, the third evaluation dimension (i.e. marketing intelligence) is probably the most abstract one. According to Büchner & Mulvenna (1998) it has been defined as the gain of additional marketing knowledge for a better understanding of the users' behaviour when using the interactive advisory system. Core objectives with respect to marketing intelligence are both the identification of coherent customer segments and outstanding preference patterns which may lead to new service developments (Woon et al. 2003; Mobasher, 2007). Thus, the evaluation approach proposed below finally employs *correspondence analysis* to visualize the interrelation between categorical characteristics of users in order to identify previously unknown patterns of interest (Backhaus, 2003). Table 1 provides a summary of the evaluated dimensions, targeted indicators and selected methods.

Table 1 Evaluation Dimensions

Dimension	Efficiency	Effectiveness	Marketing intelligence
Measure/ Indicator	Customer Satisfaction (CS) Usability	Click-through- & Conversion-Rate, Customer Retention	Identified Customer Segments, Opportunities for Product Development
Mining Method	Click-stream Sequence Analysis	Contingency Analysis	Correspondence Analysis

The choice of data mining methods mainly depends on the scaling type of the involved variables (e.g. user characteristics). In this study these were mainly nominally scaled (Hair et al. 2006).

4 Application of the Evaluation Framework

The exemplarily findings presented are based on web-usage data logged during the second half of 2006 belonging to the web-portal of an Austrian thermal resort and the corresponding interactive travel advisory system. More than 40,516 distinct user

sessions (i.e. online visits) were identified out of which approximately 2% could be unambiguously associated with an advisory session of the interactive travel advisory system. However, approximately the same percentage of sessions couldn't be related to web-log entries due to unresolved DNS-names of the accessing hosts or the use of proxy servers. Consequently, a total of 712 distinct user sessions of the interactive travel advisory system were used in the analysis.

4.1 Click-stream Sequence Analysis

In order to empirically detect the eventual friction that provoke users to quit a RS, a *click-stream sequence analysis* is proposed as a method for visualizing users' online interaction paths and showing the relative share of traffic at each node (Senecal et al., 2005). Figure 3 depicts the online conversations of users interested in wellness offerings. As can be seen from the sequence diagram, after question six users can either switch to the final page comprising the individual recommendations or can further communicate preferences to the system. Around 60% of all users reached the recommendation page. Nevertheless, it became evident that most users left the interactive advisory system already within the first three dialogue steps, which thus need to be reassessed particularly with respect to understandability and usability.

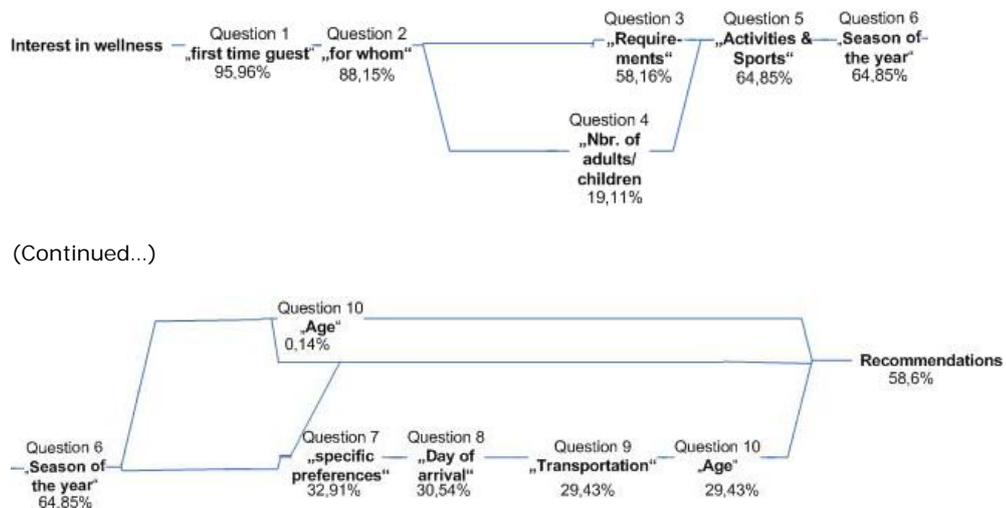


Figure 3 Click-stream Sequence Analysis

4.2 Contingency Analysis

As effectiveness of a RS has been approximated by both the click-through and conversion rate, (Zins et al. 2004) a *contingency analysis* is proposed as it allows the detection of interrelationships between categorical variables (Giudici & Castelo, 2001; Pallis et al., 2005). In doing so, this method answers the question whether the usage frequency of the interactive travel advisory system varies depending on contextual session parameters, such as access time or cultural background (i.e. preferred language). Interestingly, a significantly higher relative frequency of accesses to the RS could be observed on weekends ($\chi^2(1) = 3.848$; $p < 0.05$; see Table 2).

Table 2 Use of interactive travel advisor and access times

	Use of virtual advisor		Total in %
	Yes in %	No in %	
Access at weekend			
Yes	2,07	97,93	100
No	1,77	98,23	100

From the viewpoint of cultural groupings it was interesting to observe that Italians are much more likely to use the innovative interactive advisory tool than both German or English speaking users ($\chi^2(3) = 208.11$; $p < 0.001$; Table 3).

Table 3 Use of interactive travel advisor and cultural belonging

	Use of virtual advisor		Total in %
	Yes in %	No in %	
Language preference:			
German	1,42	98,58	100
Italian	3,14	96,86	100
English	0,46	99,54	100
Unknown	0,91	99,09	100

In a final step the hypothesis that ‘users who interact with a recommender system are more likely to enquire about accommodation availability’ was empirically tested (Zins et al., 2004). From Table 4 one can easily observe that a significantly larger share of availability requests is triggered by online visits that employed the interactive advisory system compared to the group of potential customers who did not use the system. The above hypothesis that the usage of a RS positively correlates with the number of availability requests was empirically confirmed ($\chi^2(1) = 62.87$; $p < 0.001$; see Table 4).

Table 4 Probability of booking requests and use of interactive travel advisor

Availability request	Use of virtual advisor	
	Yes in %	No in %
Yes	8,91	3,49
No	91,09	96,51
Total	100	100

4.3 Correspondence Analysis

According to Büchner & Mulvenna (1998) marketing intelligence has been defined as the gain of additional marketing knowledge for a better understanding of the users’ behaviour when using the interactive advisory system. Thus, the goal is to identify coherent customer segments and outstanding usage patterns leading to new service developments (Mobasher, 2007). The proposed empirical method of *correspondence analysis* pursues this goal to visualize relationships between a set of non-metric attributes, thus reducing data complexity and helping to group corresponding characteristics of non-metric data (Backhaus et al., 2003).

As an example, correspondence analysis was computed for the two non-metric variables ‘cultural belonging’ (i.e. *language*) and ‘usage group’. The latter construct was identified by asking users for whom the travel packages were intended. Users could choose one of the following standardized answers: *for myself*, *for myself & my partner*, *for my whole family* and *for others* (see Table 5).

Table 5 Correspondence Table

Language	For whom				
	For myself	For myself & my partner	For my whole family	For others	Active border
German	83	81	32	15	211
Italian	70	205	108	5	388
English	14	9	3	0	26
Active border	167	295	143	20	625

Table 6 shows the details of the two reduced dimensions. '*Inertia*' measures how much information of the raw data has been absorbed by the corresponding dimension (Hair et al., 2006). For instance, it can be seen that dimension 1 incorporates 91.3% of total variance, whereas dimension 2 only holds about 8.7% (Hair et al., 2006).

Table 6 Correspondence analysis

Dimension	Singular val	Inertia	Chi-Square	Sig.	Share of		Singular value for confidence	
					Inertia	Condition	Standard deviation	Correlation 2
1	,306	,094			,913	,913	,039	,236
2	,094	,009			,087	1,000	,025	
Total		,103	64,071	,000 ^a	1,000	1,000		

a. 6 Degrees of freedom

Figure 4 displays the perceptual map obtained from the correspondences between cultural belonging (i.e. language) and the usage group variable. Interestingly, the obtained findings suggest that Italian speaking users are more likely to be looking for family offers and offers for partners, while German speaking users intend to travel alone. These previously unknown item combinations are of particular interest for new (i.e. tourism) product developments that are addressed to specific tourist segments (Weiermair & Fuchs, 1999).

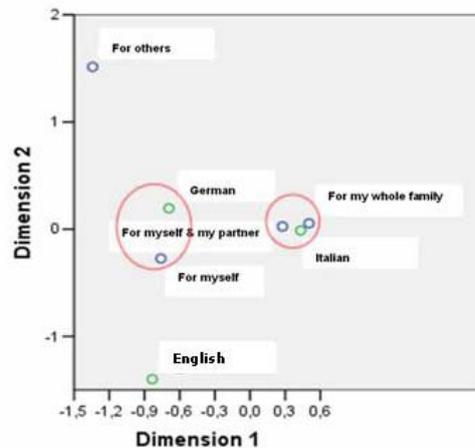


Figure 4 Perceptual Map based on Correspondence Analysis

5 Conclusions and Future Work

Based on the WEBMINER framework (Cooley et al., 1999), the paper proposed an approach for evaluating interactive recommender systems in the tourism domain. Starting from the main performance dimensions, methodological approaches for determining the value of recommender systems from an organisation view point (i.e. in terms of efficiency, effectiveness and marketing intelligence) were proposed and explored. For instance, based on a *click-stream analysis* (Senecal et al., 2005), the specific node that caused the majority of users to quit the virtual travel advisor could be identified. Furthermore, based on *contingency analysis* (Pallis et al., 2005) it was demonstrated that the usage share of the interactive travel advisor is significantly higher on weekends and is more than twice as likely to be used by Italian speaking users. Interestingly enough, the hypothesis that the usage of an interactive recommender system positively correlates with the number of availability requests was empirically confirmed. Finally, marketing intelligence was obtained by the use of *correspondence analysis* (Hair et al., 2006). For instance, of particular interest for both segment specific (i.e. tourism) product development and targeting is the notion that Italian speaking users are more likely looking for family offers and offers for a partner, while German speaking users intend to travel alone.

The discussed results should only be interpreted as selected examples to underline the usefulness of the proposed approach in evaluating interactive recommender systems in the tourism domain based on automatically collected log data. Additional insights

can, however, easily be generated by using the described methods when considering the totality of variables and their corresponding web-log data (e.g. first time guest, number of children, transportation mode, users' age and sex, preferred sports activities, etc.). Moreover, future evaluation steps will include further analytical tools which are able to handle categorical data, such as a *Chi-Squared-Automatic-Interaction-Detector* (CHAID) or *logistic regression based on binary variables* (Prinzie & Van den Poel, 2005).

References

- Backhaus, K. Erichson, B. Plinke, W. & Weiber R. (2003). *Multivariate Analysemethoden - Eine anwendungsorientierte Einführung*. Berlin, Heidelberg, New York: Springer.
- Büchner A.G. & Mulvenna M. D. (1998). *Discovering Internet Marketing Intelligence through Online Analytical Web-Usage Mining*, SIGMOD Record, vol. 27: 54-61.
- Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments, *User Modelling and User-Adapted Interaction*, vol. 12: 331-370.
- Cooley, R., Mobasher, B. & Srivastava, J. (1999). Data Preparation for Mining World Wide Web-Browsing Patterns, *Knowledge and Information Systems*, vol. 1.
- Cooley, R., Mobasher, B., & Srivastava, J. (1997). *Web-Mining: Information and Pattern Discovery on the World Wide Web*, 9th International Conference on Tools with Artificial Intelligence (ICTAI). IEEE Computer Science Washington D.C.
- Delgado J. & Davidson R. (2002). *Knowledge Bases and User Profiling in Travel and Hospitality Recommender Systems*, In K. W. Wöber, A. Frew, and M. Hitz (eds.) Information and Communication Technologies in Tourism, Innsbruck. Springer, Wien New York.
- Felfernig, A., Friedrich, G., Jannach, D. & Zanker, M. (2006). *An Integrated Environment for the Development of Knowledge-Based Recommender Applications*. International Journal of Electronic Commerce (IJEC), Special Issue on Recommender Systems, 11(2), pp. 11-34.
- Fesenmaier, D.R., Ricci, F., Schaumlechner, E., Wöber, K. & Zanella, C. (2003). *DIETORECS: Travel Advisory for Multiple Decision Styles*. In A. Frew, M. Hitz, and P. O'Connor (eds.) Information and Communication Technologies in Tourism, Helsinki. Springer, Wien New York.
- Giudici, P. & Castelo, R. (2001) Association Models for Web Mining, *Data Mining and Knowledge Discovery*, 5(3): 183-196.
- Hair, J., Black W., Babin, B., Anderson, R. & Tatham, R. (2006). *Multivariate Data Analysis*, 6th edition, New Jersey: Prentice Hall.
- Herlocker, J., Konstan, J., Terveen, L. & Riedl, J. (2004). *Evaluating Collaborative Filtering Recommender Systems*, ACM Transactions on Information Systems, vol. 22: 5-53.
- Jannach, D., Zanker, M., Jessenitschnig, M. & Seidler, O. (2007). *Developing a Conversational Travel Advisor with ADVISOR SUITE*, In M. Sigala, L. Mich, and J. Murphy (eds.) Information and Communication Technologies in Tourism, Ljubljana. Springer, Wien New York.

- Mobasher, B. (2007). *Data Mining for Web Personalization*, (eds. Brusilovsky, P. et al.) The Adaptive Web, vol. LNCS 4321, Springer: pp. 90-135.
- Pallis, G., Angelis, L. & Vakali, A. (2005) *Model-Based Cluster Analysis for Web Users Sessions*, M.-S. Hacid, Z. W. Ras, and S. Tsumoto (eds.) 15th International Symposium ISMIS 2005, Saratoga Springs, NY, Springer LNCS Vol. 3488: 219-227.
- Pan B. & Fesenmaier, D. R. (2003). *Travel Information Search on the Internet: A Preliminary Analysis*, In A. Frew, M. Hitz, and P. O'Connor (eds.) Information and Communication Technologies in Tourism, Helsinki. Springer, Wien New York.
- Prinzie, A. & Van den Poel D. (2005). Constrained optimization of data-mining problems to improve model performance. *Expert Systems with Applications*, 29(3): 630-640
- Ricci F. & Werthner, H. (2002). Case-based querying for travel planning recommendation. *Information Technology and Tourism*, vol. 4: 215-226.
- Ricci, F. (2002). Travel Recommender Systems, *IEEE Intelligent Systems*, Vol. Nov.: 55-57.
- Senecal, S., Kalczynski, P. J. & Nantel, J. (2005). Consumers' decision-making process and their online shopping behaviour: A Click-stream analysis, *Journal of Business Research*, 58(11): 1599-1608
- Weiermair, K. & Fuchs, M. (1999). The Effect of Cultural Distance on perceived Service Quality Gaps: Implications for IT-based Intercultural Communications Strategies, *Service Operations Management Association (SOMA) Conference*, Bentley College, Boston, USA, August 22-25: 127-132. Service Operations Management Association.
- Werthner, H. (2002). *Intelligent Systems in Travel and Tourism*, 18th International Joint Conference on Artificial Intelligence, IJCAI-03, Acapulco, Mexico.
- Woon, Y.-K., Ng, W.-K., Li, X. & Lu, W.-F. (2003) *Efficient Web-log mining for product development*, Proceedings of Cyberworlds, Singapore: 294-301. IEEE Computer Science Washington D.C.
- Zins, A. H., Bauernfeind, U., Missier, F., Venturini, A. & Rumetshofer, H. (2004). *An Experimental Usability Test for different Destination Recommender Systems*, In A. Frew (ed.) Information and Communication Technologies in Tourism, Cairo. Springer, Wien New York.