



Institut for Economic
Promotion,
Austrian Economic
Chamber,
Vienna, Austria



T.E.I. Thessaloniki
Greece



Department of Tourism
Management

UDC 658.8:004.738.52
Preliminary communication
Received: 17.01.2008

ANALYSING THE USERS' PERCEPTION OF WEB DESIGN QUALITY BY DATA MINING TOOLS

Vanja Bevanda
Jasmina Grzinic
Emanuel Cervar

Juraj Dobrila University of Pula, Croatia¹

Abstract: Several authors indicated the existence of different relative importance of each web site quality factor across e-business domains and between stakeholders. They also found the correlation between web site quality and e-business performance. The website with the highest quality produced the highest business performance. Therefore there is a need to constantly monitor users' behavior and their requirements in different e-business sector towards better web site design.

The aim of this research is to explore possibilities of classification data mining tools to support and automate process of discovering users' perception of web site design quality factors. In order to do that we collected 81 cases that serve as an input for inductive reasoning. We applied modified ID3 algorithm and induced the most informative attributes and rules describing users' perception of web design quality for online travel agency. The findings can provide the management and web designers with useful insights to enhance and refine their business performance.

Keywords: Web design quality, Evaluation, Online travel agency, Inductive reasoning, Data mining technique, Classification.

¹ **Vanja Bevanda**, Ph.D., Assistant Professor, **Jasmina Grzinic**, Ph.D., Assistant Professor, **Emanuel Cervar**, Graduate student, Juraj Dobrila University of Pula, Department of Economics and Tourism «Dr. Mijo Mirkovic», Pula, Croatia.

INTRODUCTION

Past development of network technology and Internet itself resulted in numerous changes not only in business, entertainment, education and society in general, but also in the process of software development. The World Wide Web is one of the most relevant driving forces in the commercial usage of the Internet. It is an important factor among the Internet applications although it still has a relatively small quantitative importance compared with the traditional transactions. The number of web sites and their size are increasing² but the web sites developments are undertaken with only limited resources (time and money) As a result, many sites are poorly designed and do not meet customers' requirements.

The field of design in new media promotes interdisciplinarity and multidisciplinary and requires research and processing of recent scientific discoveries in several connected areas (technical sciences, humanities and design). Owing to the Internet, tourist portals and networks for information exchange, individuals, agencies and tour operators have a real time control over free capacities anywhere in the world. Airlines, rent-a-car services, trip organizers and animators also take part in all that. Planning destinations has never been easier. In the modern world the web design of a tourist organization has become one of the key factors for its successful business

Much has been written about the design of websites from the human engineering, user interface, training, business, and usability perspectives. Many works are available, as well as extensive online sites with the guidelines covering design aspects such as readability, appearance, ease of navigation and searching, accuracy and reliability, etc. Existing Web page design guidelines often offer the same advice for all types of websites, regardless of their purpose (Susser& Ariga: 2006). There is no doubt that the site must be designed to fit the audience. The Web site development requires a user centered design process with permanent evaluation the evolving design against user requirements.

Another complicating factor is the fact that the WWW is an extremely dynamic environment. Site design options change very fast and must be redesign with each technological wave. We try to facilitate these activities with extensive use of data mining tools that extracts meaningful patterns and behavior models for small segment of users and do that very quickly and easy. We applied modified ID3 algorithm and induced the most informative attributes and rules describing users' perception of web design quality for online travel agency. The findings can provide the management and web designers with useful insights to enhance and refine their business performance.

1. METHODOLOGY

Besides the importance and accessibility of Internet, web design is crucial to all tourism companies. As opposed to companies operating in other activities, tourist

² In 1994, there were only 3000 web sites on Internet, and a few years later, the number was increased to 4.27 millions and 1.5 million web pages were born daily (Gonzales& Palacios: 2004).

organizations depend on seasons, having peaks in the number of users during high seasons, whether summer or winter, and significant drops after the season. In tourism users expect the locations to meet their needs and be consistent, reacting well to the geographical location, their language, technology and wishes. It is therefore differentiated through attractive flash animations, tailored web programs: system for an independent control of destinations, accommodation, arrangements and last minute offer. The design is most frequently adapted through logotypes, basic photos and number of pages and languages.

The web design quality research on tourism industry-specific issues have not reached a consensus on what makes a tourism site effective (Susser& Ariga, 2006). Survey of previous research in this field has shown that Web sites evaluation studies tend to rely on expert assessments or predetermined benchmarks and on the tangible aspects of a web sites rather than on consumers options (Park & Gretzel: 2007). Owing to communication and information technology, it is possible today to make bookings, orders and check capacities almost automatically, as well as keep track of it all through Internet. Smaller businesses (for example tourist agencies) use services of other companies designing web pages for them, while larger businesses (tourist companies, tour operators...) have their own IT department, including a web designer. Statistics say that almost 50% of tourist services users use Internet to get information or book and pay for their trip or vacation.

Web designers, especially in a tourism field, have a difficult job which requires a lot of effort, creativity, originality and which, above all, has to result in a catchy page that will attract the user of tourist services more than a page from competition. The importance of a simple but attractive look of a web page is one of the key factors to a successful business of a tourist organization (Kaplanidou& Vogt, 2006).

Unfortunately, it is not enough to simply offer a quality presentation on web pages or prepare online booking in hotels. Modern tourists want everything in a package, from all inclusive hotels, to included trips, sports and recreational contents, entertainment and night life. It all has to be united in one package and presented to the market using modern communication technologies.

In our research we consider the application of another approach from the previously done in this field. The aim is to facilitate the process of describing users' behavior using data mining tools. This tool extracts meaningful patterns and builds predictive customer-behavior models that can serve as an aid in decision making. It is a largely automated process sifting through data sets to detect useful, non-obvious, and previously unknown patterns or data trends. The emphasis is on the computer-based exploration of previously uncharted relationships (i.e., using "machine learning" methods that typically require only limited human involvement). This technology offers enormous gains in terms of performance, speed of use, and user friendliness (Magnini et al.: 2003). We performed analysis of customer behaviors when evaluating the web site design of online travel agency. Discovered findings can help managers and designers to spot trends quickly that can be applied to future web design development and adjustment.

In order to accomplish the above mentioned goal we developed a questionnaire where we identified the main factors considered as determinates of web site design quality found in previous studies. After that we performed a field study with two small groups of students who had a task to evaluate tree web sites of Croatian online travel agencies. The resulting survey serves as an input for knowledge-based system applying case based reasoning in order to find pattern in users` behaviors.

2. WEB SITE DESIGN EVALUATION FACTORS AND DATA COLLECTION

Research of tourist demand on the international tourist market proves that it has become more important for tourists how to spend an unforgettable holiday than where to spend it. Special events have to become a tourist product designed especially for tourists, which is already the case in successful tourist destinations in order to appear different on the market and create or strengthen their image and create a recognizable brand. In such cases the support of web design becomes crucial.

There are different ways in which clients reach the web page of a tourist facility, like direct access to a certain page, after which one reaches the web page through various advertisements or banners on portals. There is a high probability that some of the surfers will be interested in tourist services because of reached popularity. One of the most popular and simplest ways to reach a certain page are the search engines among which Google is the most popular. Naturally, tourist businesses have to pay a fee to have a banner on a portal, while Google charges for some key words which will make the searched web page appear among the first in the search results.

Several evaluation studies have been conducted related to tourism Web sites (Park& Gretzel, 2007) using a multitude of approaches that range from expert judgments to consumer surveys to automated evaluations by crawler technology. These studies have identified a myriad of possible factors. The past research seems to have in common a general agreement that assessing a Web site's effectiveness requires multidimensional instead of one-dimensional evaluation approaches and measures.

The number, labels, and definitions for these dimensions differ across the various studies, making it difficult to compare findings and identify factors that have consistently been used for evaluation. This situation has led to little progress in our understanding of key factors that should be included in evaluation frameworks.

Park and Gretzel, 2007, presented web sites evaluation measures from 53 tourism papers published in journals and proceedings from 1997 to 2006. They identified nine common factors that appear to form the basis of the majority of studies related to tourism sector. These factors are: Ease to use, Responsiveness, Fulfillment, Security/ Privacy, Personalization, Visual Appearance, Information quality, Trust and Interactivities.

Booking and payments for tourist arrangements on-line is an established practice abroad, while in Croatia it is all relatively new. It is predictable that the

number of services available on-line will continue to grow every year, not only in order to provide clients with more services but also to keep up with the competition.

Nevertheless, the most important tourist agencies and tour operators in the Republic of Croatia have been offering these services for a couple of years. It is still a usual practice that the on-line system enables users to book and pay for accommodation and food in tourist facilities, group and individual travels, traveling tickets for all kinds of means of transport, outings and renting of cars or boats.

Since the aim of this paper was supporting web design evaluation in sector of online travel agencies, not their web effectiveness, it was reasonable to restrict the analysis to several factors related to design. Students who were asked to evaluate selected web sites do not have any or had modest previous online shopping experience, and they visited selected sites in one session. It was not possible for them to evaluate two factors from the previous list: Security/Privacy and Trust.

Instead of these two indicators, server availability and speed of downloading appeared as very important limitation in Croatian environment. We included that in our consideration by adding a factor named "Accessibility". The final list of key web design evaluation factors with their description are given in Table 1.

Table 1. Web design evaluation factors with their description

Key Factors	Description
Visual Appearance	Site attractiveness, Aesthetics
Ease of Use	Logical structure
Fulfillment	Noticeable of special offerings
Navigability	Ease of navigation
Accessibility	Server availability and downloading speed
Personalization	Existing of advanced search function and customization
Interactivity	Use of multimedia in representing offering details
Information quality	Content presentation and currency

The subject of the study was a small group of students` population (30 students from Juraj Dobrila University in Pula) comprised of an equal number of males and females with an age range of 19–21. Each student was asked to compare three websites considering the site as a whole. The sites selected represent three commonly used Croatian online travel agencies.

Data collection took place in a controlled setting. University computer labs with one computer per subject were utilized. All subjects were given instructions and began the survey at the same time. The students were instructed to wander through each site as if they were searching for information using their regular surfing behavior. They were instructed not to complete the evaluation of the site until they had navigated

through the home page and at least three sub pages of the site. Students were supervised to minimize any discussion.

The web sites were measured using a three-point rating scale from 3 to 1, where number 1 means the best solution. An initial draft of the questionnaire was pretested. In their refinement, we restricted the number of initial factors and we added the factor "Accessibility" that appeared as very important. To avoid possible confusion with number meaning, we offer their "soft" measures description in scale from worst to the best attribute for each factor in the questionnaire. The list of factors with their values is given in table 2.

Table 2. Evaluation factors and their values

Name	Value 1	Value 2	Value 3
Web design quality (on line travel agency)	poor	average	Very well
Visual Appearance (Attractiveness)	unattractive	Like others	Strong eye appeal
Ease of use (Logical structure)	unclear	Table of content	Clear structure
Fulfillment	unobservable	available	noticeable
Navigability	hard	With effort	easy
Accessibility	slow	With patience	fast
Personalization	Non-existing	unhelpful	Easy to find
Interactivity	Not exist	Only few	Too many
Information quality	poor	adequate	rich

After removal of 2 incomplete surveys, we got evaluations from 27 students for three web sites. They showed us which web sites are perceived as prestigious and of quality, but we still do not know why they reached that conclusion.

3. DATA MINING WITH INDUCTION ALGORITHMS

The original ID3 algorithm was introduced by Quinlan (Quinlan: 1992). It is a type of machine learning in the form of decision trees. Non-incremental unsupervised learning is used. "In unsupervised learning each observation within a set is described by the same set of attributes, and this forms the nature of the input information. In the non-incremental, the system deals with finite set of observations (...) it is possible to determine the complexity of the algorithms, and this is useful in resolving real applications. However, if the authors want to consider a new observation within the

system, it is necessary to process the whole set of observations again.” (Martínez-Enríquez and Eschalada-Imaz: 1998).

The following modified ID3 algorithm is used to build a decision tree, given a set of non-categorical attributes C_1, C_2, \dots, C_m , the categorical attribute C_i , and a training set C of records. Functioning of ID3 algorithm can be described as the following pseudocode (Quinlan: 1992): Function ID3 (β : a set of non-categorical attributes, C_i : the categorical attribute, C : a training set); begin If C is empty, return a single node with value "Failure"; If C consists of records all with the same value for the categorical attribute, return a single node with that value;

If β is empty, then return a single node with as value the most frequent of the values of the categorical attribute that is found in records of C ; Let C_i be the attribute with largest Informativity $I_b(C_i, C)$ among attributes in β ; Let $\{w_i | i=1, 2, \dots, m\}$ be the values of attribute C_i ; Let $\{C_i | i=1, 2, \dots, m\}$ be the subsets of C consisting respectively of records with value w_i for attribute C_i ; Return a tree with root labeled C_i and arcs labeled a_1, a_2, \dots, a_m going respectively to the trees ID3($\beta - \{C_i\}, C_i, C_1$), ID3($\beta - \{C_i\}, C_i, C_2$), ..., ID3($\beta - \{C_i\}, C_i, C_m$); end Determining informativity (I_b) of attribute b is as follows: Let C the set of cases in node, a the benchmark, $a_1 \dots a_n$ its values, and $w_{a_1} \dots w_{a_n}$ ($\sum_i w_{a_i} = 1$) their rates in set C . Then entropy of benchmark in set C can be written: $E_C = -\sum_i w_{a_i} \log_n w_{a_i}$. Let $b_1 \dots b_n$ the values of attribute b , β is a set of them. Disjoint β into not empty subsets called $\beta_1 \dots \beta_m$. Then $\cup_i \beta_i = \beta$. Disjoint C into subsets called $C_1 \dots C_m$ being attribute b of all elements of C_i in β_i for each i . Let w_i the weight of C_i in C . ($\sum_i w_i = 1$). Then $I_b = E_C - \sum_i w_i E_{C_i}$, in words informativity is an increment of entropy resulted from disjoining $\beta_1 \dots \beta_m$. Real output of computing is $I_{b_{max}}$ of optimal selection.

4. THE USERS' PERCEPTION OF WEB DESIGN QUALITY BY INDUCTION ALGORITHM

For exploring possibilities to support analyzing the user's perception of web design quality, we use a questionnaire as an input into knowledge-based system shell called Doctus. For induction Doctus uses a ID3 algorithm described above.

Doctus uses symbolic artificial intelligence for three types of reasoning: deduction, induction and reduction. It is able to cope with tacit and implicit rules at the same time, so decision makers can clearly see, using "if... then" rules, the satisfactory solution (then and there). It reasons both deductively and inductively, so it enables the user to check on the model graph why the chosen solution is in the given case most appropriate.

With Doctus it is possible to recognize the relations between the data and selection only the needed rules to the decision maker (Baracskaï et al.,: 2002.)

The collected answers served as an input for case-based reasoning (induction algorithm). The following pictures are an excerpt from knowledge-based that consists of 81 cases (Figure 1 and 2).

Figure 1.: Doctus`s screen with Cases

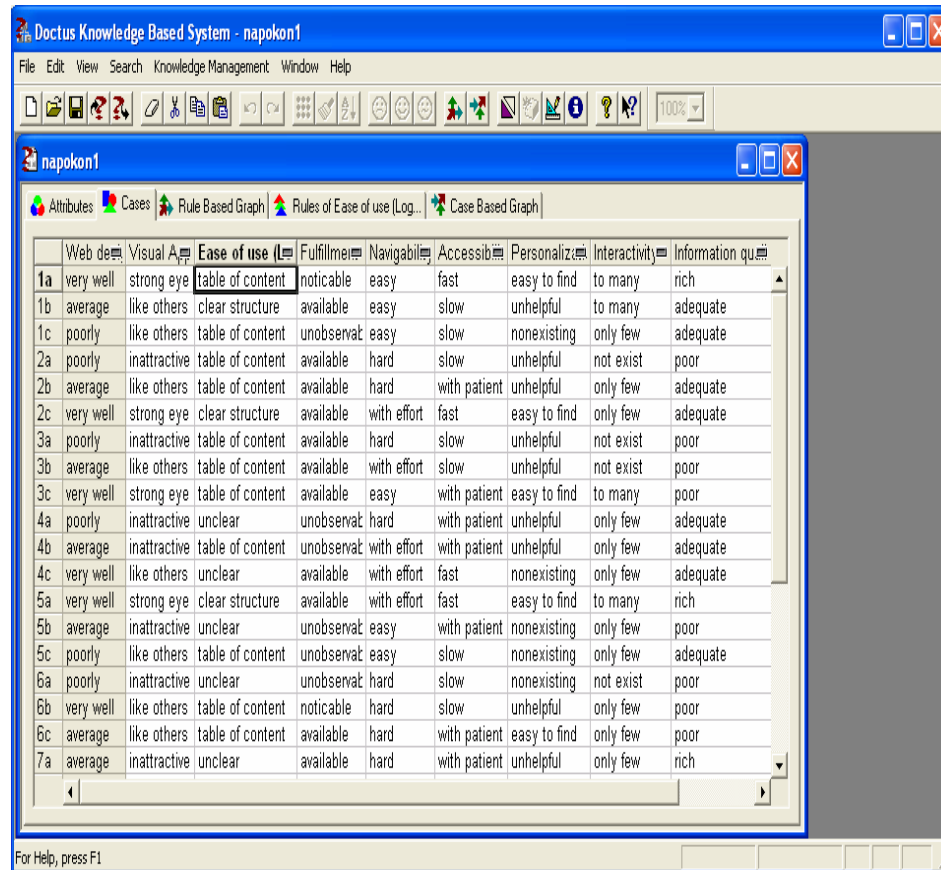
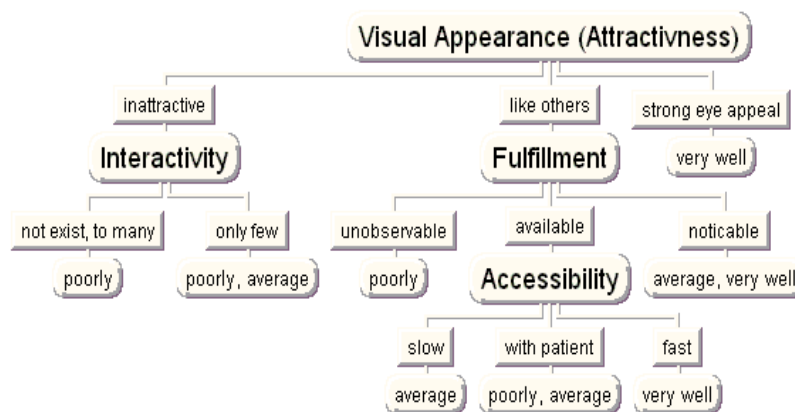


Figure 2.: Cases excerpt from knowledge base

	Webdesign	Visual Appeal	Ease of use (Clarity)	Fulfillment	Navigability	Accessibility	Personalization	Interactivity	Information content
1a	very well	strong eye appeal	table of contents	noticable	easy	fast	easy to find	to many	rich
1b	average	like others	clear structure	available	easy	slow	unhelpful	to many	adequate
1c	poorly	like others	table of contents	unobservable	easy	slow	nonexisting	only few	adequate
2a	poorly	in attractive	table of contents	available	hard	slow	unhelpful	not exist	poor
2b	average	like others	table of contents	available	hard	with patient	unhelpful	only few	adequate
2c	very well	strong eye appeal	clear structure	available	with effort	fast	easy to find	only few	adequate
3a	poorly	in attractive	table of contents	available	hard	slow	unhelpful	not exist	poor
3b	average	like others	table of contents	available	with effort	slow	unhelpful	not exist	poor
3c	very well	strong eye appeal	table of contents	available	easy	with patient	easy to find	to many	poor
4a	poorly	in attractive	unclear	unobservable	hard	with patient	unhelpful	only few	adequate
4b	average	in attractive	table of contents	unobservable	with effort	with patient	unhelpful	only few	adequate
4c	very well	like others	unclear	available	with effort	fast	nonexisting	only few	adequate
5a	very well	strong eye appeal	clear structure	available	with effort	fast	easy to find	to many	rich
5b	average	in attractive	unclear	unobservable	easy	with patient	nonexisting	only few	poor
5c	poorly	like others	table of contents	unobservable	easy	slow	nonexisting	only few	adequate
6a	poorly	in attractive	unclear	unobservable	hard	slow	nonexisting	not exist	poor
6b	very well	like others	table of contents	noticable	hard	slow	unhelpful	only few	poor
6c	average	like others	table of contents	available	hard	with patient	easy to find	only few	poor
7a	average	in attractive	unclear	available	hard	with patient	unhelpful	only few	rich
7b	poorly	like others	table of contents	available	with effort	with patient	unhelpful	not exist	poor
7c	poorly	in attractive	unclear	unobservable	hard	with patient	unhelpful	not exist	poor
8a	average	like others	table of contents	noticable	easy	fast	easy to find	to many	adequate
8b	poorly	in attractive	unclear	available	with effort	with patient	unhelpful	not exist	adequate
8c	very well	strong eye appeal	clear structure	unobservable	hard	with patient	nonexisting	to many	rich
9a	very well	strong eye appeal	clear structure	unobservable	with effort	with patient	unhelpful	not exist	poor
9b	average	like others	table of contents	noticable	easy	fast	easy to find	only few	adequate
9c	poorly	in attractive	unclear	noticable	hard	slow	nonexisting	to many	rich
10a	very well	strong eye appeal	clear structure	available	easy	with patient	easy to find	to many	adequate
10b	poorly	in attractive	unclear	unobservable	hard	fast	unhelpful	not exist	rich
10c	average	like others	table of contents	noticable	with effort	slow	nonexisting	only few	poor

After inductive reasoning we got the resulting decision tree. Figure 3 presents the decision tree describing the users' (students') perception of web design quality factors. It can be described by the rules as follow: "If *Visual Appearance (Attractivness)* has value *"strong eye appeal"* in surveys, than *Quality of web site design* has value *"very well"* or "If *Visual Appearance (Attractivness)* is *"like others"* and *"Fulfilment"* is *"unobservable"*, than *Quality of web site design* is *"poorly"*.

Figure 3.: Decision tree describing users' perception of web design quality factors for three online travel agencies



For practitioners, the implications of these results reinforce what many site designers have tried to articulate: make it simple. Adopting a minimalistic approach to the design of the home page with eye-catching but appropriate graphics and categories that draw the web surfer further into the site appears to be more effective.

Web design should not result in information overload. The goal, rather, should be to give access to the information web surfers desire in the most expedient way possible. Hence, the design goal should be access not abundance. Simplicity of design should be a major consideration as it not only makes the site more appealing, but also makes it far faster to load. Web surfers are not a patient group. It is not, therefore, surprising that slow loading sites are a major frustration and turnoff for web surfers. Another prerequisite is to make the website attractive. A website with an identity will appeal to web surfers, differentiate the company and make the site more memorable.

RESULTS AND DISCUSSION

One of the constant themes of site development guides is that the site must be designed to fit the audience. This means that the design fitting the user will continue to

be the main prerequisite for tourist agencies to be working online. In an environment offering wide choice, the users will not remain at sites that do not meet their needs. The user simply abandons a site with a confusing interface or a site which is too slow.

Many sites have multiple segments to which they cater. If preferences are distinctly different among different segments, different options might be made available to different groups. What must be determined is how much flexibility must be built into website content to satisfy the increasing diversity of users.

A major complicating factor to all of this is the fact that the web is an extremely dynamic environment. Site design options change with each technological wave. By providing site designers with a better idea of how to facilitate interacting through the cognitive landscape of the web, simplicity may be more effectively accomplished.

Research results confirm that the Web will not satisfy all the needs of its users – it will be enough for a site to gather a good number of visitors or offer content no other site is offering. From the web designers perspective, one of the early activities in web site development process are to define the business objectives, the intended context of use and key scenarios of use. This helps prioritize design and provides a focus for evaluation. Web site development requires a user centered design process with permanent evaluation of the evolving design against user requirements.

One of the greatest advantages of the kind of business described in the paper is the huge amount of content. The key factor in web designing is marking safe pathways through information so that such an amount of content does not represent a problem.

Guides – professionals will filter the information, evaluate the content, act as guides and even guide groups on tours. We facilitate these activities with extensive use of data mining tools that extract meaningful patterns and behavior models for a small segment of users and do that very quickly and easily.

REFERENCES

- Baracskai, Z.; Bevanda, V., Dorfler, V. (2002) Intelligent Customer Relationship Management (iCRM) by eFlow Intelligent Portal, 13th International Conference on Information and Intelligent Systems, Conference Proceedings, Varazdin, 173-180.
- Gonzales, M. F. J.; Palacianos, T. M. B. (2004) Quantitative evaluation of commercial web sites: an empirical study of Spanish firms, *International Journal of Information Management* 24, 313-328, Available at: www.elsevier.com/locate/ijinfomgt, Accessed: 2007-10-22.
- Kaplanidou, K. & Vogt, C. (2006) A Structural Analysis of Destination Travel Intentions as a Function of Web Sites Features, *Journal of Travel research* 45 (2), 204-216.
- Lee, Y.; Kozar, K. A. (2006) Investigating the effect of website quality on e-business: An analytic hierarchy process (AHP) approach, *Decision Support Systems* 42, 1383-1401, Available at: www.Elsevier.com/locate/dss, Accessed: 2007-10-22.
- Magnini, V. P.; Honeycutt, E. D. Jr. and Hodge S.K. (2003) Data Mining for Hotel Firms: Use and Limitations, *Cornell Hotel and Restaurant Administration Quarterly* 2003; 44; 94, <http://cqx.sagepub.com>, Accessed: 2007-11-07.

- Martínez-Enríquez, A.M. & Eschalada-Imaz, G. (1998). The revision of inductive learning theory within incomplete and imprecise observations. *Expert Systems with Application*, 15, pp. 357-366. Available from: <http://elsevier.com/locate/eswa>. Accessed: 2000-06-03.
- Morgan, N.J. et al. (2001) Consumers, travel and technology: A bright future for the Web or television shopping? *Journal of Vacation Marketing*, 7(2): 110 - 124. o.j:n
- Park, Y. A.; Gretzel, U. (2007) Success Factors for Destination Marketing Web Sites: A Qualitative Meta-Analysis, *Journal of Travel Research*, 46, 46. Available at: <http://jtr.sagepub.com/cgi/content/abstract/46/1/46>, Accessed: 2007-11-05.
- Perkins, E. (2000) Online Travel, *Microsoft Press; 1 edition*.
- Quinlan, J. R. (1992) C 4.5: Programs for Machine Learning. *Morgan Kaufman*, ISBN 1558602380.
- R. Law and J. Wong, J. & R. Law (2003) Successful Factors For A Travel Web Site: Perceptions Of Online Purchasers In Hong Kong, *Journal of Hospitality & Tourism Research*, 27(1): 118 - 124.
- Tania C. Lang, T. C.O (2000) The effect of the Internet on travel consumer purchasing behaviour and implications for travel agencies, *Journal of Vacation Marketing*, Vol. 6, No. 4, 368-385, (2000).