The knowledge destination – a customer information-based destination management information system

Wolfram Höpken [b/a],
Matthias Fuchs [a],
Dimitri Keil [b], and
Maria Lexhagen[a]

[a] European Tourism Research Institute (ETOUR)
Mid-Sweden University, Sweden
{matthias.fuchs, maria.lexhagen}@miun.se

[b] Business Informatics Group
University of Applied Sciences Ravensburg-Weingarten, Germany
wolfram@hoepken.org
dimitri.keil@gmx.de

Abstract

Huge amounts of customer-based data, such as tourists' website navigation, transaction and survey data are available in tourism destinations, however, remain largely unused (Pyo, 2002). This paper presents the concept of a knowledge-based destination management information system (DMIS) that supports value creation through enhanced decision making. Information is extracted from heterogeneous data sources of the Swedish tourism destination of Åre and is categorized in explicit feedback (e.g. survey data) and implicit information traces (e.g. navigation and transaction data). Methods of business intelligence are applied to retrieve interesting data patterns, thus, to generate knowledge in the form of empirically validated models. The paper deduces new insights about the applicability of data mining techniques and related models at tourist destinations depending on the type of tourism data and concrete problem characteristics at hand (Pick & Schell, 2002).

Keywords: management information systems, business intelligence, data mining, tourism knowledge destination

1 Introduction

The competitiveness of tourism destinations depends largely on how information needs of stakeholders can be satisfied through ICT-based infrastructures and services (Buhalis, 2006). However, although huge amounts of customer-based data are widespread in tourism destinations (e.g. webservers store tourists' website navigation, data bases save transaction and survey data) these valuable knowledge sources typically remain unused. Thus, managerial competences and organisational learning in tourism destinations could be significantly enhanced by applying methods of *business intelligence* (Min *et al*. 2002; Pyo *et al*., 2002; Sambamurthy & Subramani,

R. Law et al. (eds.), *Information and Communication Technologies in Tourism 2011*

2005). The latter method offers reliable, up-to-date and strategically relevant information about tourists' travel motives and service expectations, channel use and related conversion rates, booking trends, estimations about the quality of service experience and value-added per guest segment (Pyo, 2005).

Through the generation, management and access of relevant information, the knowledge level of tourism destination stakeholders can be increased and information asymmetries be reduced. Knowledge relevant to tourism suppliers (i.e. information about tourists and destination resources) fosters market cultivation and destination competitiveness is strengthened by enhancing service effectiveness using destination resources in a more sustainable way. However, it is less the knowledge base existing at any time per se, than a firm's ability to effectively apply (and learn from) existing knowledge to create new knowledge and to take strategic action that forms the basis for achieving competitive advantage. Thus, the major challenge of knowledge management is to make individual knowledge about customers, products, processes, competitors or business partners available and meaningful to others (Pyo *et al*., 2002). This makes also clear why ICT and methods of business intelligence are playing a crucial role in effectuating the knowledge-based view of the firm (Grant, 1996) when enhancing large-scale intra and inter-firm knowledge exchange.

The objective of the paper is to address the above deficiencies by conceptualizing, prototypically developing and validating a destination management information system (DMIS) that supports value creation through enhanced supplier interaction and decision making. Methods of business intelligence are applied to the Swedish tourism destination of Åre to retrieve relevant and previously unidentified knowledge from customer-based data. New insights about the applicability of data mining techniques in tourism destinations depending on the type of data and problem characteristics will be achieved. As a prerequisite for competence development, the proposed approach supports knowledge generation by enhancing knowledge transfer and absorption processes at the level of the destination management organisation (DMO) and small and medium enterprises (SMEs).

The paper is structured as follows. Section 2 presents the concept of a knowledge destination framework and describes the different aspects of customer-based knowledge generation and knowledge application in the form of a knowledge-based DMIS. Section 3 describes how the knowledge destination framework has been validated at the Swedish tourism destination of Åre. Section 4 summarizes the results and gained insights and provides an outlook on future research activities.

2 Knowledge Destination Framework

The conceptual foundation of the proposed DMIS is the *knowledge destination framework*. Accordingly, knowledge activities deal with extracting information from different customer and supplier-based sources as well as with generating relevant knowledge and applying it in the form of intelligent services for customers or suppliers (i.e. destination stakeholders). Thus, the knowledge destination framework distinguishes between a *knowledge creation* and a *knowledge application layer*.

The *knowledge creation layer* extracts information from heterogeneous data sources and makes destination-specific knowledge available to tourists and destination suppliers. On the customer side, content is generated by tourists through feedback mechanisms providing sources of *structured* (guest surveys, evaluation platforms, etc.) and *un-structured* data (free-text, blogs, etc.). Moreover, implicit knowledge can be made explicit by visualizing tourists' information traces (web search, navigation) through online application tracking and web-mining (Liu, 2008; Pitman *et al*., 2010). Furthermore, valuable knowledge about buying behaviour is generated through mining transaction (booking) data. Finally, tourists' mobility behaviour may be traced by GPS/WLAN-based position tracking (Zanker *et al*., 2010). On the supplier side, knowledge about products, processes, competitors and strategic partners is extractible from existing data sources or websites, e.g. in the form of destination profiles or availability information (Ritchie & Ritchie, 2002; Pyo, 2005).

The *knowledge application* layer provides knowledge-based services for customers as well as destination suppliers and stakeholders. An emerging area of knowledge-based customer services lies in the field of location-based services (Berger *et al*., 2003), comprising services that automatically push context-sensitive messages to tourists, recommend destination services, support dynamic device adaptation and are sensitive to the users' context (Höpken *et al*., 2008). A second prominent application area lies in the field of community services typically used to learn from prior consumption experiences of others (Xiang & Gretzel, 2010). Examples are community sites (e.g. LonelyPlanet), review sites (e.g. TripAdvisor), blogs and blog aggregators (e.g. blogspot.com), social networking sites (e.g. Facebook) and media sharing sites (e.g. YouTube, Flickr). Finally, multi-modal human computer interfaces supporting interactive search services are crucial to enhance knowledge creation in tourism.

By contrast, knowledge-based supplier services mainly fall into the category of tourism-related business intelligence applications (Cho & Leung, 2002; Olmeda & Sheldon, 2002). Explorative data analyses, online analytical processing (OLAP) and data mining (i.e. predictive analytics, clustering, association rules and classification) allow the de-centralized (ad-hoc) generation, management and access of strategically relevant knowledge to the DMO as well as private and public destination suppliers (Fuchs & Höpken, 2009). Thus, crucial management functions that support strategic decision making are provided (e.g. business performance management, forecasting, multi-channel and online community management, dynamic pricing).

The paper proposes a knowledge-based DMIS where the knowledge application layer focuses on supplier-based knowledge application. Although, for knowledge generation both supplier and customer-based information is of principal relevance, the proposed approach restricts itself to customer-based knowledge generation.

2.1 Customer-based knowledge generation

Fig. 1 illustrates all components of the knowledge destination framework. The knowledge generation layer is comprised of the different customer-based data sources and components for data extraction, data warehousing and data mining described in more detail in the following sections.

Data sources: Since the uptake of CRS/GDS in the 1960s' a major part of tourism transactions are handled electronically. With the rapid growth of the WWW this portion further increased, and nowadays customers leave electronic footprints during all travel-related activities, like searching and trip planning, reservation & booking, service consumption (e.g. using mobile services and GPS/WLAN-based position tracking or loyalty programmes, like customer cards) and, finally, post-trip activities in community web sites.

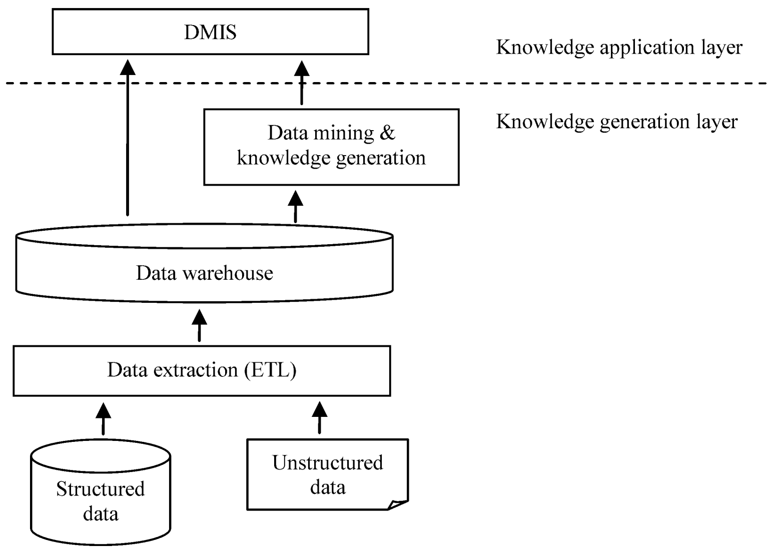

Fig. 1. Knowledge destination framework architecture

Thus, huge volumes of data on customer transactions (e.g. customer inquiries, bookings, payments processing), customer needs and behaviour are typically stored by different stakeholders of a tourism destination. Main added-value is not just a comprehensive collection of data from different sources but especially their combination to generate new knowledge, e.g. the continuous analysis of customer behaviour in all trip phases. Table 1 provides a systematization of data sources to be considered by the proposed knowledge destination framework.

Table 1. Potential data sources of the knowledge destination framework

Explicit tourists' feedback provided knowingly and intentionally	**Implicit tourists' information traces provided unknowingly and unintentionally**
- *Structured data*: e.g. online and offline guest surveys, ratings from web 2.0 applications, user profiles from web applications and online communities, etc. - *Unstructured data*: free text from E-mails and web 2.0 applications (e.g. blogs, e-comments/reviews), rich content (e.g. YouTube.com), etc.	- *Navigation data*: search behaviour on web sites and online portals, community sites, etc. - *Transaction data*: online requests, reservation and booking data, payment, etc. - *Tracking data*: GPS/WLAN-based coverage of tourists' spatial movements - *Observation data*: gathered in a laboratory context or through market observation

Data extraction: Obviously, different data sources require different techniques for the extraction of relevant information, dependent on the data format at hand. *Structured data*, regardless of whether being the results of customer surveys or customer transactions, typically exist in quite differing formats. Thus, a key challenge in the field of information extraction is the integration of heterogeneous data sources by using semantic, linguistic or constraint-based techniques of information integration (Dell' Erba, 2005). By contrast, *unstructured* (or semi-structured) *data* is considered in the form of free text or *html*-documents (e.g. written feedback, blogs or other community-based content). In the case of *html*-documents information extraction is achieved through wrappers, either created manually-based on static patterns or (semi-) automatically generated by means of (un-)supervised learning methods (Liu, 2008). In the case of unstructured data in the form of free text, information extraction is achieved by methods of text mining based on statistical language models (Manning, & Schütz, 2001) or approaches of natural language processing (Jurafsky & Martin, 2000). In order to conform to data privacy aspects, critical personal data (i.e. name, exact address) has to either be left out or obfuscated (Hastie *et al*. 2009).

Data warehousing: At the core of the knowledge destination is a central data warehouse (DW) that embraces all data relevant to tourism stakeholders (Cho & Leung, 2002). Thus, heterogeneous data from different data sources are mapped into a homogeneous data format and stored in a central DW. Only through this harmonisation process is it possible to carry out a destination wide and all-stakeholder encompassing analysis approach. Based on a tourism ontology and methods of semantic annotation and transformation, individual data sources are, first, transformed into a central data model and, finally, into a dimensional structure typical for a DW.

Data mining and knowledge generation: Based on the data collected in the DW, the main task is to generate relevant knowledge for destination suppliers and the DMO. By employing methods of data mining (i.e. techniques of machine learning and artificial intelligence) interesting patterns and relationships in the data are detected and knowledge will be provided in the form of validated models (e.g. clustering models, classification models or association rules). The presentation of these models and the underlying data (in the case of exploratory data analysis or online analytical processing – OLAP) clearly rests on the *knowledge application layer* (section 0). Beside the data itself stored in the data warehouse, the data mining models are the central output of the knowledge generation layer, either used as input to static reports or, and this is the foremost case in the context of the knowledge destination framework, being interactively visualized within DMIS, described next.

2.2 Supplier-based knowledge application – a destination management information system (DMIS)

As suggested by literature, knowledge relevant to strategic decision making in tourism is subsumed as a) knowledge about *market cultivation* (e.g. how to use destination resources to attract most valuable customers and to provide information for effective marketing, etc.), and b) knowledge relevant for *destination management & development* (e.g. facilities to avoid congestion, environmental protection, development of product-market combinations for valuable segments, event management, training, private-public partnerships, etc. (Wang & Russo, 2007; Bornhorst *et al*., 2010). Accordingly, customer-based knowledge creation is achieved through customer segmentation and targeting, service performance evaluation and by measuring the degree of marketing efficiency (Ritchie & Ritchie, 2002; Pyo *et al*., 2002; Cho & Leung, 2002). Thus, data collected, stored, analyzed and visualized in the DMIS include demographic/geographic and psychographic characteristics, buying motives, price sensitivity, brand perception and loyalty as well as information and product consumption patterns, respectively. Table 2 provides a not exhaustive list of indicators relevant for a DMIS. Columns represent basic data mining methods while the rows display sources of customer-based data available in tourist destinations. The indicators have been deduced from the quoted literature.

Table 2. DMIS indicators

Explorative analyses (OLAP)	**Clustering**	**Association rules / classification / prediction**
Navigation based indicators • Web-navigation & channel use (page frequency, view time, path length, click-streams)	***Navigation based indicators*** • Web usage-based clusters (key-word, session or transaction-based) • Webpage-based clusters	***Navigation based indicators*** • Sequential navigational patterns • Visitor forecast based on online search volume
Transaction based indicators • Sales shift per accomm. type/ sending country/guest type • Booking patterns per tourist activity/ tourism service • Conversion rate per guest type / sending country	***Transaction based indicators*** • Valuable segments based on o demographics & consumption behaviour o mobility behaviour in the destination • Conversion rate/ segment	***Transaction based indicators*** • Cross-sales (market basket analysis) • Cancellation Behaviour • Occupancy trends & length of stay/ sending country/ guest type
Feedback-based indicators • Guest satisfaction, value for money assessment & loyalty per sending country/guest type • User generated content: Ø rating, % positive reviews	***Feedback-based indicators*** • Valuable segments based on brand comprehension, value for money assessment, satisfaction & loyalty • Social interaction clusters	***Feedback-based indicators*** • Social network dynamics based on UGC o lead users

3 Validation of DMIS Indicators

The knowledge destination approach has been prototypically implemented for the Swedish tourism destination of Åre in order to validate the DMIS indicators. Although including also small-sized suppliers, Åre is characterized by two large scale companies, Ski Star and Holiday Club. Next to Åre's DMO (Åreföretagarna), the listed companies are the partners of an EU funded project from which the proposed approach emerged (*acknowledgements*). The goal was to apply the idea of business intelligence at the level of tourism destinations by mining data bases usually not further analyzed through methods of artificial intelligence (Pyo *et al*., 2002).

3.1 Data sources, data extraction and data warehousing

The prototype provides an exemplary instantiation of all components of the knowledge destination framework (Fig. 1). Compared to all possible data sources (section 2.1) the prototype is restricted to *implicit* tourists' information traces (i.e. navigation and transaction data) constituting the most comprehensively available data sources for the destination of Åre, at the moment. More concretely, *navigation data* (i.e. web server log-file data) has been integrated from the platforms www.visitare.se, www.are360.se and www.holidayclub.se, as well as *transaction data* from customer data bases of Holiday Club and Ski Star.

Log-file data are typically provided in a standardized format, e.g. the *Common Logfile Format, Extended Logfile Format* or similar formats (www.w3.org/TR/WD-logfile.html). Thus, log-file-based navigation data from different sources are rather easily extracted and integrated into a homogeneous structure. Based on the user's IP address, navigation data is further enhanced by information on location (country, city, etc.) and the Internet provider (www.ip2location.com). For data preparation, in a first step, irrelevant page views caused by search engine robots or internal requests by developers have been removed. Secondly, page views have been mapped to meaningful page categories, corresponding to modules or user actions, relevant for further analysis. Finally, single page views have been grouped to sessions based on a simple time heuristic (maximal time between two consecutive page views within a session) and session-specific attributes, like session length and clicks per session, have been generated (Liu, 2008).

Compared to log-files, data extraction and data preparation for company-specific data bases containing transaction data is much more complex and cumbersome and may count for up to 60% of the overall data mining effort (Larose, 2005). Transforming heterogeneous data base structures into a homogeneous supplier-embracing structure for transaction data has been outside the scope of this prototype, thus, transaction data from different sources have been analysed separately. But, even when looking at single data sources, complex and normalized data base structures have to be transformed into flat tables to subsequently serve as input for data mining activities and data understanding and extraction still turned out to be complex and time-consuming. Data preparation has mainly dealt with removing irrelevant or highly correlated attributes or handling missing values and outliers. All steps of data extraction and preparation have been executed with the data mining tool *RapidMiner* (www.rapid-i.com). The resulting data are available as *RapidMiner*-specific data files (example sets) as well as data base tables that all together constitute the central DW serving either as input for the *data mining and knowledge generation* task or directly for visualization purposes by the DMIS (fig. 1).

3.2 Data mining & knowledge generation

Following the DMIS indicators (table 1) and based on available data (section 3.1), data mining analyses have been executed, making use of explorative analyses, clustering, association rules and classification, thereby covering navigation- and transaction-based indicators (Hastie *et al*., 2009).

In the area of navigation-based indicators by explorative analysis typical web-navigation and channel use metrics, like page frequency, view time, session length, etc. have been generated. Table 3 and fig. 2 show examples of web-navigation metrics aggregated along the time dimension as well as clicks grouped per webpage category.

Table 3. Web-navigation metrics – sessions and clicks per time unit

Month (session)	Count(session)	Avg.(session_lenght)	Sum(clicks_per_session)	Avg.(clicks_per_session)
Dec	7210	2.042	2867	3.970
Jan	25445	2.190	104253	4.097
Feb	32315	2.173	129250	4.000
Week in year	Count(session)	Avg.(session_length)	Sum(clicks_per_session)	Avg.(clicks_per_session)
52	3380	2.005	13235	3.916
1	5314	2.114	21570	4.059
2	6222	2.045	23846	3.833
Day in week	Count(session)	Avg.(session_lenght)	Sum(clicks_per_session)	Avg.(clicks_per_session)
Mon	10645	2.177	43696	4.105
Tue	10144	2.147	40679	4.010
Wed	10037	2.427	43446	4.329

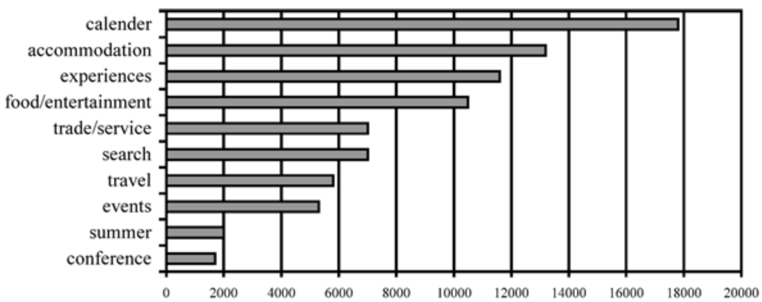

Fig. 2. Web-navigation metrics - clicks per webpage category

For the area of navigation-based clustering, table 4 shows a *k-means clustering* model based on web-usage characteristics, like visited pages (URIs mapped to meaningful page categories), clicks per session and session length. Based on cluster quality and interpretability, a five-cluster model has been chosen. Comparing these clusters enables a clear differentiation between *bookers* (cl. 1) and *lookers* (cl. 3 and cl. 4). In cluster 1, users enter the website with a clear booking intention searching for the right accommodation. In cluster 4 users spend most of the time looking at panorama pictures and searching for accommodation information, but seldom enter the booking page. Thus, either they have not decided on the overall destination yet or have no concrete purchase intention. Appropriate (cross-selling) offers could increase the conversion rate of cluster 4 users.

Table 4. Cluster model - segmentation of website visitors

Cluster Model		**Attribute**	**cluster_0**	**Attribute**	**cluster_1**
		ownership	0.115	accommodation	1.512
Cluster 0:	3239 items	accommodation	0.141	accomm_hotel_room	1.227
Cluster 1:	564 items	whatson_start	0.121	accomm_appartment	0.922
Cluster 2:	1078 items	booking	0.107	acomm_booking	0.621
Cluster 3:	736 items	directions	0.103	offers	0.583
Cluster 4:	82 items	activities_summer	0.091	booking	0.498
Total number:	5699 items	restaurant_bar	0.091	accomm_suite	0.389
		restaurant_bar_sports	0.076	offers_private	0.259
Attribute	**cluster_2**	**Attribute**	**cluster3**	**Attribute**	**cluster_4**
offers	1.308	pool_saunaworld	1.037	panorama_pictures	8.378
offers_privat	0.663	pool_sauna_pool	0.887	ownership_pan_pict	1.866
accommodation	0.319	spa	0.764	media	0.841
activities_summer	0.218	spa_treatments	0.516	accommodation	0.780
booking	0.186	pool_sauna_sauna	0.420	accomm_appartment	0.634
ownership	0.161	activities	0.295	pool_saunaworld	0.439
accomm_package	0.153	accommodation	0.245	offers	0.427
activities	0.118	offers	0.236	pool_sauna_pool	0.402
Cluster	**Avg(session_length_minutes)**		**Avg.(clicks_per_session)**		
cluster_0	0.869		2.471		
cluster_1	5.128		10.454		
cluster_2	2.407		4.995		
cluster_3	3.618		7.844		
cluster_4	9.020		18.224		

In the area of navigation-based indicators through association rules / classification, association rules have been developed to identify navigational patterns and understand the booking behaviour of website visitors through session characteristics. Table 5 shows association rules based upon the *a-priori algorithm* (Larose, 2005, p. 180f) representing associations between session characteristics (premises) and "visited the booking sub-site" (conclusion). These rules identify session characteristics influencing the likelihood to visit the booking page. More precisely, heavy users with sessions longer than 5 minutes and more than 11 clicks visit the booking page 3.5 times more often than average users. Interestingly enough, visiting the booking page is more likely during evening/night hours and during weekdays than during daytime and/or on weekends.

Table 5. Association rules - booking behaviour

Premises	**Conclusion**	**Lift**
session_length=heavy user(5-x minutes), clicks/session=heavy user(11-x clicks)	accomm_booking=visited	3.576
clicks_per_session=heavy user(11-x clicks)	accomm_booking=visited	3.105
origin=key country(SWEDEN), cliks_per_session=heavy user(11-x clicks)	accomm_booking=visited	3.063
day in week=during week(Mo-Fr), clicks_per_session=heavy user(11-x clicks)	accomm_booking=visited	2.994
hour in day=during day(5-17h), clicks_per_session=heavy user(11-x clicks)	accomm_booking=visited	2.886
session_length=heavy user(5-x minutes)	accomm_booking=visited	2.741
day in week=during week(Mo-Fr), session_length=heavy user(5-x minutes)	accomm_booking=visited	2.591
hour in day=during night(18-4h)	accomm_booking=visited	1.192
day in week=during week(Mo-Fr), hour in day=during night (18-4h)	accomm_booking=visited	1.184
origin=key country(SWEDEN), hour in day=during night(18-4h)	accomm_booking=visited	1.178

In the area of transaction-based indicators by explorative analysis booking and sales patterns have been analysed. Table 6 shows aggregated information on customer segments and room categories (capacity) in relation to number of bookings, number of visiting days, visiting persons and booked rooms. The figures confirm Åre's focus on domestic customers which should be viewed in relation to marketing activities towards foreign visitors. Fig. 3 shows the relationship between date of booking and date of arrival and unveils a strong correlation between booking behaviour and type of season (high season like December to April; low season like May, September or October). This might be caused by the fact that in low season customers typically postpone their booking since they expect the destination not to be fully booked.

Table 6. Bookings per customer segment

Segment	Bookings	Days	Avg. days	Persons	Avg. persons	Rooms	Avg. rooms
individual_domestic	6887	16168	2.348	20394	2.961	7946	1.154
individual_foreign	2660	5238	1.969	8740	3.286	3178	1.195
company_domestic	1486	3224	2.196	8060	5.490	5104	3.477
TS guests	1450	3287	2.267	4994	3.444	1460	1.007
company_foreign	1433	3030	2.114	8927	6.230	4916	3.437
recurrent private	1051	2223	2.115	3530	3.359	1104	1.050

Transaction-based clustering focuses on identifying valuable customer segments based on demographics (e.g. gender, age, country, customer type) and consumption behaviour (e.g. product type, lodging period, booking source, booking status) (Hastie *et al*., 2009). Table 7 shows the results of a cluster analysis based on a customer relationship database of customers of Åre. Cluster 0 is the typical private leisure traveller, booking accommodation for a longer period by phone and web, but cancelling quite often. Cluster 1 is the typical business traveller booking accommodation for a short period, interestingly, mostly by phone. Cluster 2 represents private customers booking ski rental mostly by phone.

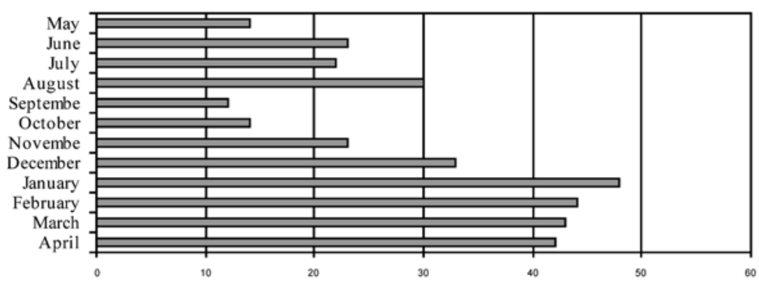

Fig. 3. Duration between date of booking and arrival (in days)

As these customers do not book any accommodation and do not register for a regular use of the system, it can be assumed that cluster 2 contains customers either have their own accommodation or book on a private basis. Questions arise as to why customers in clusters 1 and 2 do not book via the Internet and how the online platform could be improved to attract such customer segments. This is especially important for cluster 2, since customers booking no accommodation at all or on a private basis will not appear in any destination statistic as long as they don't buy additional services. Thus, attracting these customers to use the online platform would greatly increase their visibility for destination management.

Table 7. Customer segments by demographics and consumption behaviour

Cluster 0 (31343)	Cluster 1 (29774)	Cluster 2 (38883)
Lodging period = week Customer type = private Booking source = phone + web Booking status = cancelled Product type = lodging Profile: private customer, booking a one week lodging via web and phone, often cancelling	Lodging period = short week / weekend Customer type = company / N/A Booking source = phone Product type = lodging + other Profile: short trip company customer, booking lodging + other products via phone	Lodging period = N/A Customer status = (non-registered) customer Customer type = private Booking source = phone Product type = ski rental Profile: not registered private customer, booking ski rental via phone

Finally, in the area of transaction-based indicators through association rules / classification, customers' cancellation behaviour has been analysed using *decision tree* based classification models (with target attribute booking status and input attributes departure/arrival/booking date, gender, age, country, customer type, product type, lodging period and booking source) (Larose, 2005, p. 107f).

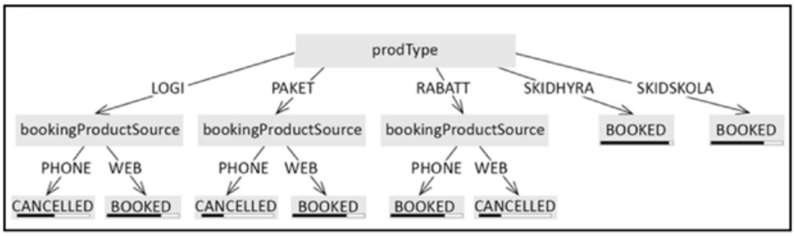

Fig. 4. Decision tree for cancellation behaviour

Fig. 4 clearly shows that the cancellation behaviour differs for different product types with ski equipment (*skidhyra)* and ski school (*skiskola*) showing a low cancellation rate, whereas accommodation (*logi*) and package (*paket*) are showing high cancellation rates. Interestingly, for accommodation and package the cancellation rate depends on the booking source and is particularly high when booked via phone and quite low when booked via the Internet. Overall, the presented decision tree model can predict the cancellation behaviour with an accuracy of 92.9%.

4 Conclusion and Future Work

The proposed DMIS fully builds on destination data stored in a DW and knowledge generated from that data in form of data mining models. Moreover, the data and data mining models are interactively visualized in an appropriate form for destination managers. Customer-based data is displayed by methods of exploratory data analysis and validated models, like *decision trees* (e.g. showing the factors determining cancellation behaviour) or *cluster models* (e.g. describing segments of website users) and are visualized by simple but effective visualization techniques (e.g. simplified decision rules and tree structures). Most importantly, however, the gained insights clearly demonstrate that data typically available within tourism destinations, like in the case of the Swedish destination of Åre, enable the computation of navigation- and transaction-based DMIS indicators. Thus, the proposed knowledge destination framework and the DMIS indicators constitute an appropriate and validated base to apply business intelligence and knowledge generation methods at the level of tourism destinations.

Compared to potential data sources listed in Table 1, only a limited range of different data sources has been considered, so far. Thus, one important future activity is to include all possible types of data sources, especially feedback-based data, like survey data, web 2.0 ratings or user profiles as well as unstructured data, especially in the form of free text from web 2.0 applications. Additionally, knowledge generation has to be extended to the area of supplier-based knowledge generation, thus, generating knowledge about products, processes, competitors and strategic partners. Similarly, the knowledge application layer will be extended by the area of customer-based knowledge application and generated knowledge will particularly serve as input for recommender systems or social media aggregators. Thus, also customer demand is satisfied in an intelligent and collaborative manner and the destination offer will continuously be improved, based on customer needs and market knowledge.

References

Berger, S., Lehmann, H. & Lehner, F. (2003). Location-based Services in the Tourism Industry. *Information Technology & Tourism* 5(4): 243-256.

Bornhorst, T., Ritchie, J.R. & Sheehan, L. (2010). Determinants for DMO & Destination Success: An Empirical Examination, *Tourism Management* 31: 572-589.

Buhalis, D. (2006). The Impact of ICT on Tourism Competition. In A. Paptheodorou, (Ed.), *Corporate Rivalry and Market Power.* IB Tauris, London: 143-171.

Cho, V. & Leung, P. (2002). Knowledge Discovery Techniques in Database Marketing for the Tourism Industry, *J. of Quality Assurance in Hospitality & Tourism* 3(3): 109-131.

Dell'Erba, M., Fodor, O., Höpken, W. & Werthner, H. (2005). Exploiting Semantic Web Technologies for harmonizing e-Markets. *Information Tech. & Tourism* 7(3/4): 201-220.

Fuchs, M. & Höpken, W. (2009). Data Mining im Tourismus, *Praxis der Wirtschaftsinformatik* 270 (12): 73-81.

Grant, R.M. (1996). Toward a knowledge-based View of the Firm. *Strategic Management Journal* 17(2): 109-122.

Gretzel, U. & Fesenmaier, D.R. (2004). Implementing a Knowledge-based Tourism Marketing Information System: the Illinois Tourism Network, *Information Technology & Tourism*, 6: 245-255.

Hastie, T., Tibshirani, R. & Friedman, J. (2009). *The Elements of Statistical Learning – Data Mining, Inference and Prediction* (2nd. Ed.). Springer, NY.

Höpken, W., Scheuringer, M., Linke, D. & Fuchs, M. (2008). Context-based Adaptation of ubiquitous Web Applications in Tourism. In P. O'Connor, W. Höpken, & U. Gretzel

(Eds.), *Information and Communication Technologies in Tourism* (pp.533-544). NY: Springer.
Jurafsky, D. & Martin, J.H. (2000). *Speech and Language Processing.* Prentice-Hall, NJ.
Larose, D.T.(2005). *Discovering knowledge in data.* John Wiley & Sons, New Jersey, 2005
Liu, B. (2008). *Web Data Mining* (2nd ed.), Springer, New York
Manning, C.D. & Schütz, H. (2001). *Foundations of Statistical Natural Language Processing*. MIT, Cambridge.
Min, H., Min., H. & Emam, A (2002). A Data Mining Approach to develop the Profile of Hotel Customers. *Int. Journal of Contemporary Hospitality Management* 14(6): 274-285.
Olmeda, I. & Sheldon, P.J. (2002). Data Mining Techniques and Applications for Tourism Internet Marketing, *Journal of Travel & Tourism Marketing* 11(2/3): 1-20.
Pick, R.A. & Schell, G. P. (2002). An Architecture for Knowledge Management Featuring Meta-Knowledge. In C. Holsapple, V. Jacob, & H.G. Rao (Eds.) *Business Modelling – Multidisciplinary Approaches.* Boston: Kluwer.
Pitman, A., Zanker, M., Fuchs, M. & Lexhagen, M. (2010). Web Usage Mining in Tourism. In: Gretzel, U, Law, R. & Fuchs, M. (eds.), *Information and Communication Technologies in Tourism 2010* (pp. 393-403). NY: Springer.
Pyo, S. (2005). Knowledge-map for Tourist Destinations. *Tourism Management* 26(4): 583-594.
Pyo, S., Uysal, M. & Chang, H. (2002). Knowledge Discovery in Databases for Tourist Destinations. *Journal of Travel Research* 40(4): 396-403.
Ritchie, R.J.B. & Ritchie J.R.B. (2002). A Framework for an Industry supported Destination Marketing Information System. *Tourism Management* 23, 439-454.
Sambamurthy, V. & Subramani, M. (2005). Information Technologies and Knowledge Management, *Management Information Systems Quarterly* 29(1):1-7.
Wang, Y. & Russo, S.M. (2007). Conceptualizing and evaluating the Functions of Destination Marketing Systems. *Journal of Vacation Marketing* 13(3): 187-203.
Xiang, Z. & Gretzel, U. (2010). Role of Social Media in Online Travel Information Search. *Tourism Management* 31: 179-188.
Zanker, M, Jessenitschnig, M. & Fuchs, M. (2010): Automated Semantic Annotation of Tourism Resources based on Geo-Spatial Data, *Information Technology and Tourism* 11 (4): 341-354

Acknowledgements

This research project has been financed by the EU *Structural Fund* objective 2 project no. 39736, Sweden. The authors would like to thank the CEOs Lars-Börje Eriksson (*Åreföretagarna*), Niclas Sjögren Berg (*Ski Star Åre*) and Marketing Assistant Hans Ericson (*Holiday Club Åre*) for their excellent cooperation.