Advanced Technologies in e-Tourism

PAVEL ČECH, VLADIMÍR BUREŠ
Department of Information Technologies
Faculty of Informatics and Management
University of Hradec Královo
Rokitanskeho 62, Hradec Králové 500 03
CZECH REPUBLIC
pavel.cech@uhk.cz; vladimir.bures@uhk.cz

Abstract: - The aim of this paper is to present and provide a critical review of the advanced technologies that can be used in e-tourism. The main attention is paid to technologies that deal with some kind of knowledge representation. The paper tries to indentify major drawbacks that prevent wide spread use of these technologies in the tourism sector and thus points to some of the possible research areas. The paper focuses particularly on data mining, semantic web and multi-agent technologies.

Key-Words: - data mining, semantic web, ontology, multi-agent technologies, tourism

1 Introduction
Recent advancements in ICT enable the rise of a digital economy. In digital economy there is a convergence of ICT and other technologies that fosters the information exchange, increases demand for digital products and services and the transforms processes to a digital form [21]. The digital as well as physical products equipped with chips and networking capabilities are being delivered through digital infrastructure. The digital economy brings new needs and opportunities of how to conduct business in a more productive way [37]. Thus businesses are stimulated to adopt various e-activities such as e-business, e-commerce, e-marketing, e-invoicing and concepts such as e-identity or e-services. In the tourism sector the e-tourism is now becoming of a lively debate among academics and practitioners. However, the use of advanced technologies for information and knowledge processes are rather sporadic. The aim of this paper is to present and provide a critical review of the advanced technologies that can be used in e-tourism. The main focus is on technologies that deal with some kind of knowledge representation. The paper tries to indentify major drawbacks that prevent wide spread use of these technologies in the tourism sector and thus points to some of the possible research areas.

2 Overview of e-tourism
The computer and internet mediated business environment denoted as e-business is now widely encompassed in most of the economic sectors. The expansion of e-business is further propelled with increasing number of internet users. It is estimated that the number of internet users grew in between the years 2000 – 2008 of about 350 % world wide totaling around 1.5 billion internet users today [30]. According to the W3C 95 % of internet users use the internet for searching travel related information [7]. The tourism sector is hence not an exception in ICT and e-business adoption. Tourism as such is a dynamically expanding economic sector, with a considerable impact on basic economic indicators and plays an important role in forming the regions. Travel exports outperform in terms of money spent any other commercial service sector i.e. sectors with trade in telecommunication services, financial services, transport and others [41]. The necessity to consider tourism as an information intensive industry has been already pointed out in various studies mentioned for example in [15]. The importance of information arises from the fact that travel related products and services are based on imagination, trust and believe and therefore cannot be experienced before. Further, the traveler is often put in more or less unknown situations and conditions and thus in the state of information deficiency so it is natural to ask for information to lower the uncertainty [20]. Hence, the penetration of ICT to travel industry is fairly obvious and preceded. However, the effort to gather information on one hand and provide on the other together with massive storage capacity and broadband connectivity devaluate information due information overload phenomena. The traditional approaches to data filtering and information retrieval seem not to reflect the current information volume and needs of humans [24]. As expressed by John Naisbitt “We drown in information and thirst for knowledge”. This is because ICT systems use relatively fixed structures for working with information that are encoded in applications and
cannot be changed easily to address new information challenges connected with knowledge management [5]. Thus there is a space for applications that use various representations of knowledge and imitate the human approaches to data filtering by employing sophisticated approaches from artificial intelligence and machine learning. The tourism sector calls for these technologies since the products relate to experiences and emotions very tightly.

3 Problem Formulation

Despite the fact that the tourism sector was typically quick in an effort to address the tourism specific needs with ICT thus incorporating e-sales, e-marketing and also for e-invoicing, tourism companies is lagging behind in employing the advanced technologies for an internal process optimization as compared to companies in other sectors [9, 11]. Advanced technologies can be successfully utilized in this sense and therefore pose a potential waiting to be harnessed. This, however, is not a straight forward process. There are drawbacks that hamper the adoption of these advanced technologies in the tourism sector. The reason is because there are only marginal improvements of these systems to handle the knowledge. The paper critically reviews the often researched technologies that were already relatively successfully introduced in other sectors. The goal is to present a realistic picture of possibilities of these advanced technologies. Apart from specific problems that will be also presented the particular attention is paid to how these systems incorporate and/or handle knowledge. The authors of this paper consider the following aspects to qualify the technology to handle the knowledge appropriately: a) knowledge representation; b) flexibility to adapt to new situations; c) synthesis of knowledge out of data; d) application of knowledge in performing given tasks. The following sections of the paper focus particularly on data mining, semantic web and multi-agent technologies.

4 Data Mining

The huge amounts of data that are generated in e-tourism applications and stored in data warehouses represent a potential for finding interesting hidden information. Knowledge discovery in databases in general and data mining in particular constitute a set of principles, approaches, tools and techniques that can be used to extract valuable information from the vast amounts of stored data. Data mining refers to the process of discovering potentially useful patterns, relations, correlations and rules which were unknown and beyond human manual capacity [25]. Data mining is an interdisciplinary field that utilizes advanced technologies for storing the necessary amounts of data; statistics that comes with approaches of how to describe and explain specific characteristics found in data; and artificial intelligence applying intelligent approaches and machine learning techniques to recognized potentially useful patterns in data. The commonly used approaches of data analysis in data mining include data description, clustering also known as segmentation, classification, regression analysis, or dependency analysis.

The application of data mining in the tourism is reportedly rare [9]. The use of data mining is mainly for the purposes of the demand forecasting. With respect to the effort to find more accurate forecasting approaches the data mining is regarded as promising alternative to traditional econometric and time series models. The data mining features are already incorporated in some of the major tools e.g. Sabre has been using data mining approach for spatial data analysis, customer ranking, selling opportunities assessment and time series modeling [34]; Amadeus offers statistical processing and trend analysis in the amount of reservation viewed from the perspectives of hotel, markets, time, provision, etc.[2]; similarly data mining can help regional authorities to understand patterns of how a certain locality is being visited thus providing a foundation for sustainable control of a tourism in that given locality.

Besides demand forecasting there exist also applications of data mining approaches in various implementations of tourism and hospitality recommender systems. Recommender systems usually utilize a sort of market basket analysis i.e. association rules discovery to understand the user buying/using patterns and build the user profile accordingly and then employ content filtering or collaborative filtering techniques to recommend products and services that match with the pattern in the user preferences [26]. Recommender systems appear to be of a great importance in the tourism sector since people often consider the opinion of others, relatively impartial travelers, as being of a good guidance. Out of the characteristics that made the product or service amenable for recommendation presented in [36] the tourism sector reflect especially on the appreciation of users to be provided with a few alternatives out of which the user can make the final selection itself.

The approaches to recommendation can be split into two categories. The statistical approach that is based on aggregating user behavior and/or previous user behavior and knowledge based approach that is built on an explicit model of user behavior. Both types of approaches have some drawbacks. The statistical approach requires to have a fairly large number of cases and the knowledge based depends on intensive
engineering [6]. In practice these two categories of approaches are combined by enhancing the statistical approach with domain specific heuristics and also often attempting to store the user preferences using some kind of knowledge representation. In particular the approaches range from the nearest neighbor algorithm to Bayesian networks also including clustering, classification, or association rules [23, 32].

The particular applications of recommender systems may be seen in sectors similar to tourism in terms of product characteristics i.e. primarily intangible experience based product or service. For example Amazon is suggesting other books to buy based on collaborative filtering principles [32]. Other examples include NewsDude for recommending news articles [3] or MovieLens for recommending movies [29]. The travel and hospitality recommender systems include for example the restaurant recommendation that in the new city guides the traveler to the restaurant that is similar to those that s/he liked before [4]. Personalized recommendation of travel destination was also offered by the DIETORECS project. The project tried to develop a recommendation in the form of a travel plan that accommodates user preferences in a destination selection process [12].

The implementation of the recommender system poses several issues that need to be addressed. First, there is a question of how to build the user profile and correspondingly how to represent the knowledge about the user preferences and also the description of the products so that it is possible to match user preferences and products. In these systems the knowledge is formalized in the form of a structured dimensional model [26] called in dataware housing the star schema. The problem is that the so called knowledge, or better, represented knowledge about the case is squeezed into a feature based vector. Hence much of the knowledge has to be reduced by this approach despite the fact that this approach attempts to view a problems from different perspectives e.g. user, product, context, evaluation, etc. The features that form the dimensions have to be selected during the development and are usually hard coded so it is not possible or difficult to incorporate new features during the use. The particular values in these features has to be manually assigned or extracted from the description of these products and services since there are mostly intangible in nature. This process is prone to be subjective for one person but does not reflect the subjective evaluation of some one else.

Next, the final decision whether a certain product should be recommended or not is based on a computed index sometimes called index of confidence or coefficient (e.g. Person’s) further still purging the knowledge. Further, it generalizes e.g. if the price is high then the restaurant should not be recommended. The user might however be willing to accept a higher price if there is a favorite meal in that restaurant. These implicit combinations among various features are hidden or hard to track and thus often not used in recommending. These combinations can only be revealed, first, if the user can find out that there was such a case e.g. a restaurant which s/he might like and was not recommended and, second, the user can provide feedback to the system.

Other problems arise in building the user profiles with preferences. Basically there exist two approaches of building the user profile. In the first approach data can be collected either explicitly by making the user to manually fill some fields or answer questions. The second approach attempts to gather the preferences implicitly by observing the user behavior. The effort that is required to assign accurate ratings is a major obstacle for explicit approach. This bottleneck can be leveraged by inferring rating from observable behavior to the system [18]. There are however several limitations either and these limitations are ranging from technical to legal and personal issues. Recommendation in the tourism sector put some of the problems into a different perspective. The sparsity problem inherent in collaborative filtering approach is reduced by the fact that the products and services would be rated at least by someone or put down from the market [31].

5 Semantic Web

5.1 Basic Principles

Semantic web is the concept of adding meaning to the data that are presented on the web. The absence of semantics on the web hampers the use of the data for other purposes than human reading. By adding the meaning the concept of the semantic web enables better processing, cooperation and sharing of the web content among applications. Associated semantic descriptions can be used by automated reasoners to infer implicit meaning and relations which would provide for more contextually relevant information [1]. In tourism the web is used heavily as a communication channel for marketing products and services and also for transactions such as booking, buying etc. creating a plethora of data and information [33]. Hence contextually supported searching for information is very useful for travelers since they often search without a clear idea what they are trying to find so the mere keyword search is not very efficient. By contextual search the results might include information which are not in an explicit relation but still are relevant and on the other hand the information that does not match based on the given context are excluded (e.g. if we are searching for a trip with sightseeing in Barcelona then information about sightseeing and
Barcelona and not with trip related information will be excluded). Further reasons for semantic web in the tourism sector are: a) tourism is a cross-sectoral industry; b) the overwhelming majority of companies are small and micro in size; c) product is an aggregation of other products and services which by their description are not trivial [13].

One problem of semantic web in general demonstrated in tourism in particular is that no site contains sufficient information even for the human reader. It means that adding the meaning by describing the data, using annotation for example, cannot do the trick to achieve the vision of the semantic web. This raises the question whether there is a lack of semantics or lack of content [17]. It can be argued however that the lack of content on one site in this case is not a problem but a generic feature/limitation of the web since no site i.e. node in the network can contain all information or cannot be fully exhaustive. There is a web of such sites mutually interlinked creating the content space and if there is enough semantics, the automated processing can help in gathering the appropriate content. Thus it stresses the importance for the semantic web even further since it can be used for inter-site content integration. This can be in the tourism sector exercised in promoting products and services which are information intensive. The final product in tourism is usually a bundle or a package of several components each of which often presented on a different websites. To match the user preferences the packages are composed dynamically while reflecting several constraints. In this regards, the automated semantic search can improve the efficiency of the dynamic packaging and personalization of the final product to fit the customer needs and expectations [7].

5.2 Ontologies
The meaning should be ensured by providing the data with a meta-level that describes the data and satisfies the condition of well-defined meaning. This meta-level is represented as some form of ontology.

Ontology is based on logical model that enables concepts to be defined as well as described in some declarative formalism such as language. Mostly the ontology consists of description of concepts and other representational terms so that to process and handle by applications with some specific purpose.

The ontology can serve various purposes. Today, the development of ontology is mostly guided by the following intentions [38]:

- promoting common understanding among a group of people (experts, engineers, etc.)
- communication support among computer systems enabling interoperability between heterogeneous applications with different data sources
- tool for knowledge engineers for the purpose of a design and development of knowledge based application

Thus automated reasoners working above the ontology can make inferences based on general rules of logic reflecting the specified relations. Ontology is however difficult to use for storing domain specific or user specific heuristics. Ontology is a fruitful as a knowledge engineering tool for mapping concepts and sharing these among applications. Nevertheless, there is the engineering effort necessary to categorize the concepts and to specify the relations. The ontology does not give any clue to operational issues meaning how concepts in the ontology can be processed, prioritized or rules applied.

In the ontology knowledge is stored in the declarative taxonomical form specifying concepts and relations in a particular domain. Usually the ontology is arranged along the “is-a” or generalization relations creating the hierarchies of concepts from general to more specific. Specifying other relations enables to make simple inferences by which concepts can be implicitly related or categorized (e.g. if the Nice is a seaside resort and the trip contains the stay in Nice then the trip can be automatically categorized as seaside trip). The search engine can make use of this implicit relation or categorization better reflection on the user intentions (continuing from the previous example: the user might be searching for a seaside trip and the search engine will return the trip to Nice among the results even though the description of the trip does not contain the explicit information that it is a seaside trip).

So far there exist a numerous ontology implementations. Nearly all ontology implementations are based on similar structure, however, the terminology differs significantly. This causes confusion that is most dominant when there is a need to port ontology from one ontology management system into another. Similarly there is rarely the case that in a certain domain a standardized ontology exists. If there is one the problem might be with its completeness and further extensibility. Usually ontology are specific purpose projects and not shared to the needed extend. So the presented cooperation and sharing of data and information among applications would require implementing some concept mapping schemas.

In the tourism domain there has been built several ontologies along several projects covering various areas of traveling. However, ontologies are sometimes focused on covering the general terms but not the specifics that is needed to describe and differentiate the characteristics of destinations, products and services in detail [33]. As
particular example of efforts to promote common understanding of terms the World Tourism Organization (WTO) comes with the thesaurus on Tourism & Leisure Activities as an authoritative source of terminology [40]. Nonetheless, this is not an ontology per se since, first, it is not formalized in some machine processable language such as Web Ontology Language (OWL), and second, it only defines terms but relations are not given explicitly making only little space for implicit categorization and inferring other implicit relations. Harmonise ontology is another attempt to provide a meta-level description of information about traveling. It has been developed during the Harmonise project and is now central to HarmoNet - Harmonisation Network for the Exchange of Travel and Tourism Information which is the successor of the Harmonise project and the HarmonTEN project [16]. The Harmonise ontology declares about 50 concepts from the tourism domain. The relations are defined only as compositions among concepts and slots missing again the potential for further reasoning above the ontology. Yet another ontology has been built by the working group DERI in cooperation with the Semantics Technology Institute at the University of Innsbruck. The ontology contains around 20 concepts mainly to cover the accommodation in a certain location and corresponding activities that can be enjoyed in that location. With the goal to integrate and ensure interoperability of major GDS systems the OTA (Open Travel Alliance) has been established. It is an organization uniting many key subjects in the tourism industry such as airlines, car rental, rail, and tour companies and Global Distribution Systems. Under the auspices of OTA the specification for information exchange has bee developed. OTA specification defines common messages that are used among trading partners. These messages include: “availability checking, booking, rental, reservation, reservation canceling and modifying, query services for service details and quality, insurance quote request for all of the hotel, airline, vehicle sectors as well as the commission exchange services, and the statistical information services.” [10]. Similar to WTO thesauri OTA specification is not an ontology. However it is a de facto standard and is used for ensuring interoperability as is the case of the semantic web.

6 Multi-agent technologies
The dispersed information sources over the entire web often not containing the complete information is a challenge on one hand for semantic enrichment of data so that, on the other hand, the automated processing with the help of various agent technologies is possible. Agent is an entity constructed to continuously and autonomously fulfill certain tasks in order to achieve a given goal while in an environment which is perceived using sensors and influenced using actuators. Agent approach should not be limited to artificial agents or artificial intelligent agents only. Agent can be a human traveler who is driving a car to a given destination and is sensing the environment such as the speed, navigation, landscape, etc. and is influencing the environment by changing the speed, direction, etc. This perspective makes it possible to examine the various agents either human or non-human in their concerted effort to achieve the given goal. The field of ambient intelligence is using this approach in which the border between human and the technology is being blurred in some respect. In the tourism industry the attention is being paid predominantly to non-human information agents however.

Multi-agent technologies are finding way how to make the number of autonomous systems to cooperate so that the mutual objective is being achieved. In multi-agent systems the resources such as knowledge, information, capability and expertise distributed so that individual agent is not able to accomplish the task alone or at least can accomplish its task more effectively [19]. The other advantages of multi-agent collaboration are of the same kind as of human teamwork e.g. shortening time to find solution since agents can work in parallel and asynchronously; higher flexibility because new agents can be added to satisfy new conditions; higher robustness since agents are relatively independent and a failure of one agent will not threaten the whole system; reuse of agents – special purpose agents can be reused in other systems.

Due to the inherent distribution of information resources and the composition of products and services in the tourism domain the multi-agent systems constitute a vital technology. This is further supported by the fact that in tourism the information resources are proprietary to a particular organization which often have incoherent goals and create autonomous entities with sometimes little concerns about integration and insufficient communication interfaces. In such cases the use of the multi-agent systems is particularly well-suited [35]. Thus the tourism industry can profit from the interest in the problem of locating information sources, accessing, filtering, and integrating information to solve a problem with the help of agents that has attracted much attention in academia and in practice. Last but not least is the trend to personalization which is evident not only in tourism but also elsewhere. The multi-agent approach addresses this trend by reserving agents to take into account personal preferences and constraints and also by learning these preferences by observing the user behavior [22].

The major issues in this effort is how to orchestrate the agents so that they do not overlap in their tasks and
how to design the architecture of the knowledge base above which the agents operate so that it is distributed for easy use by individual agents and the integrity is maintained at the same time when knowledge and expertise is shared among agents [19, 26]. Other issues concerns the granularity of agents (coarse vs. fine), heterogeneity of agents (redundant vs. specialized), methods of distributing control (benevolent vs. competitive, team vs. hierarchical, static vs. shifting roles), and communication possibilities (blackboard vs. messages, low-level vs. high-level, content) [8].

In this paper the focus is especially on how the knowledge is processed and represented. The agents operate on the knowledge base which is perceived slightly differently then in recommendation projects where primarily descriptive information of products or users is stored. Agent’s knowledge base usually contains also operational knowledge of the agent such as tasks that have been performed based on user requirements in the form of a query and solution that represent information to solve the task. This can be extended with information about the completed tasks, most frequent tasks, time to complete certain tasks if the time is an issue for certain problems etc. [26]. The knowledge base can contain private and shared elements [19]. The consistency of the knowledge base (e.g. in between private and shared parts) can be ensured by the Truth Maintenance System (TMS) which revises the so called beliefs i.e. asserted facts in case when new information should be added into the knowledge base. Hence the knowledge can be easier extended through the synthesis of the shared parts, though a certain structure of the information stored is given. The operational aspect of the knowledge base reflects the nature of knowledge in which it is perceived as the precondition to action [39] and hence the knowledge is applied to perform goal oriented tasks.

All these issues highlighted above make the development of multi-agent system rather complicated. The necessity to encode the agent with flexible behavior requires thorough engineering effort. The deployment of multi-agent system is also intricate due to the complexity of the networked processing of task and information making it hard to indentify inconsistencies, check and find errors and provide necessary corrections.

Despite the non trivial nature of the multi-agent systems, there exist several applications that utilize the multi-agent approach. In [22] the vision of a so called advocate agent is introduced. The advocate agent is claimed to facilitate product or merchant search or decision making of a user by learning human preferences and constraints, mine semantic content, identify new opportunities for action, propose them and finally transact them [22]. Another agent application for searching among the semantic services is presented in [14]. The system provides support in specific situations such as flooding in which the information from various sources needs to be integrate in a timely manner and served for various users. The agents in the system monitor information from sources with refugee movements, meteorological and hydrological information and generate reports for involved authorities containing different subsets of information.

There are also applications aimed at supporting the traveling. The system that arranges the meetings while taking into consideration personal agendas and flight schedules is described in [28]. The need for interconnectedness of schedules and dependencies of corresponding travel arrangements and the necessity to reconcile incompatibilities cast this problem to be formulated as constraint satisfaction problem. The system is composed of agents gathering information and for task completion. The information gathering agents are Agenda agent that accesses personal agendas of users and Flight scheduler agent that checks the availability of flights. The Planner Agent and Scheduler agent are responsible for accomplishing the task of arranging the meeting.

The application of the multi-agent systems in the tourism domain is described for example in [27] where the case based reasoning approach is combined with the metaphor from swarm intelligence to help the negotiation process among agents and recommend the best travel to the user. The agents are given the user preferences and then they search the whole case base for the best match. Once the best matching case is found the agent(s) “dances” to that case are recruiting other agents. The case with the highest number of agents dancing around is recommended to the user The disadvantage of that approach is that the case base is central to all agents.

### 7 Conclusion

The use of the advanced technologies in the tourism sector is still in its infancy. It can be assumed however that it will extend in the future proportionately to the increasing volume of data and information and also to changing information needs of ordinary users especially towards customization and personalization.

The implementation of advanced technologies into the company is closely connected with financial investments into the information knowledge infrastructure and corresponding processes. Thus the widespread use of the advanced technologies depends on how it can be integrated in the existing infrastructure and how the company is able to allocate money for necessary changes. The problem of the tourism sector is that it consists mostly of middle or small sized companies that usually cannot afford to allocate sufficient budget for these projects. Therefore, the progress in use of these technologies is slower compared to other sectors.
References


[34] B. Smith, "Behind the Scenes : Leveraging Technology to Produce Results, Empowering solutions symposium," Los Angeles, 2002.


