The Advent of Semantic Web in Tourism Information Systems

Dimitris N. Kanellopoulos

Technological Educational Institute of Patras

22. October 2005

Online at https://mpra.ub.uni-muenchen.de/25397/
MPRA Paper No. 25397, posted 24. September 2010 15:17 UTC
THE ADVENT OF SEMANTIC WEB IN TOURISM INFORMATION SYSTEMS

Dimitris N. Kanellopoulos  
Technological Educational Institute of Patras

The tourism industry depends on complex value creation chains involving a large number of participants that change frequently and rapidly. In addition, the products of tourism industry are complex and they will perish if they are not sold in time. For these reasons, the ideal tourism information systems require a lot of flexibility of underlying systems. Moreover, they comprise accurate access to any tourism service that provide, and they are usable by corporate and private customers alike. The management and interoperation of semantically diverse tourism information systems are facilitated by Semantic Web technology that provides methods and standards, which allow accurate access to information as well as flexibility to comply with needs of tourism information system users and administrators. This paper considers state-of-the art issues (ontologies, semantic modelling and querying, semantic portals and semantic-based e-markets) concerning the exploitation of the semantic web technologies and applications in tourism information systems.

Keywords: tourism information systems; semantic web; ontologies; semantic web services

INTRODUCTION

Nowadays, customers in tourism are increasingly less loyal, take more frequent vacations of shorter duration and take less time between choosing and consuming tourism products (Werthner & Klein, 1999). Not to mention the fact that the travel industry was one of the earliest electronic commerce adopters (Werthner & Ricci, 2004). Travel industry is one of the most important kinds of commerce through the Web, representing almost 40% of all global electronic commerce and one that most reflects the impact that this technology can have in the business process itself (Carroll, 2002). Information dissemination and exchanges are the key backbones of the travel industry, and applying to this industry the semantic web technology is a very promising approach. The Semantic Web enables better machine processing of tourism information on the
Web, by structuring web documents in such a way that they become understandable by machines (Berners-Lee et al., 2001). The semantic web allows tourism content to become aware of it. This awareness allows users and software agents (viz. Internet-based programs that are created to act autonomously) to query and infer knowledge from tourism information quickly and automatically. Semantic web technologies will influence the next generation of tourism information systems by providing interoperability, reusability and shareability among them (Maedche & Staab, 2002).

Currently, the travel industry has developed open specifications messages, based on extensible Markup Language (XML), to ensure that messages can flow between industry segments as easily as within (Dell’Erba et al., 2002). For example, the Open Travel Alliance (OTA, 2004) is an organization pioneering the development and use of specifications that support e-business among all segments of the travel industry. The cumulative effort of various teams, individuals, associations, companies, and international organizations, including air, car, cruise, rail, hotel, travel agencies, tour operators and technology providers, has produced a fairly complete set of XML-based specifications for the travel industry.

The OTA adopted the Web Services model that provides the travel industry with an ideal platform to confront the difficult problem of data heterogeneity. This problem occurs because various tourism information systems use different meta-data (viz. objective data about data) for representing their tourism resources. Web services technology is a collection of standards that allows tourism web server applications to “talk” to each other over the Internet. These standards are:

- **XML** (http://www.w3.org/XML/) for driving web application services (viz. XML schema is used in requests and replies).
- **The SOAP** (Simple Object Access Protocol: http://www.w3.org/TR/soap) provides a means of messaging between a service provider and a service requestor.
- **WSDL** (Web Services Description Language: http://www.w3.org/TR/wsd/) as the service description language.
- **UDDI** (Universal Description, Discovery and Integration: http://www.uddi.org/) as the service discovery protocol to find other tourism web applications.

Semantics can be used in the discovery, composition and monitoring of web services (Ouzzani 2004). Semantically isolated pieces of tourism information can be connected, and the user can find tourism information sources more easily, while individual tourism offers can be achieved.
In this paper we consider state-of-the-art issues concerning the exploitation of semantic web technologies and applications in tourism information systems. The rest of this paper is organized as follows. Section 2 presents tourism ontologies and section 3 discusses applications of them. Section 4 describes tourism information semantic modelling and querying. Section 5 presents semantic portals and semantic web services, while section 6 considers e-markets and intelligent software agents that exploit semantics. Section 7 discusses the sociological implications of the semantic web in the destination management organizations context. Lastly, section 8 concludes the paper with some interesting remarks.

TOURISM ONTOLOGIES

The goal of the Semantic Web initiative is to provide an open infrastructure for intelligent software agents and web services. This infrastructure is based on formal domain models (ontologies) that are linked to each other on the Web. The domain model of an ontology can be taken as a unifying structure for giving information in a common representation and semantics. An ontology comprises the classes of entities, relations between entities and the axioms which apply to the entities of that domain (Mizoguchi, 2004). Through the use of metadata organized in numerous interrelated ontologies, tourism information can be tagged with descriptors that facilitate its retrieval, analysis, processing and reconfiguration. In addition, ontologies can offer a promising infrastructure to cope with heterogeneous representations of tourism web resources. Data heterogeneity can be solved, if semantic reconciliation with respect to the domain ontology is provided between the different tourism information systems. For the tourism industry, the development of ontologies is fundamental to allow machine-supported tourism-related data interpretation and integration. A brief presentation of tourism ontologies follows.

The TOVE project (http://www.utoronto.ca/tove/toveont.html) resulted in several e-business ontologies, which specify various aspects of a tourism enterprise. The modelling of an enterprise was guided by different sets of constraints on the processes executed inside an enterprise. Core tourism ontologies will contain knowledge about the domain of travel and tourism for developing intelligent tourism information systems. In the OnTour project, a working group at the Digital Enterprise Research Institute (DERI) deployed the e-Tourism ontology (Prantner, 2005) using OWL (Web Ontology Language). The e-Tourism ontology (http://e-
Dimitris N. Kanellopoulos

tourism.deri.at/ont/) was based on an international standard: the “Thesaurus on Tourism & Leisure Activities” (viz. a very extensive collection of terms related to the area of tourism) of the World Tourism Organization (WTO, 2002). This ontology describes the domain of tourism and it focuses on accommodation and activities.

Mondeca’s tourism ontology (http://www.mondeca.com) defines tourism concepts based on the WTO thesaurus. These concepts include terms for tourism object profiling, tourism and cultural objects, tourism packages and tourism multimedia content.

Another research group developed a comprehensive and precise reference ontology named COTRIN (Comprehensive Ontology for the Travel Industry) (Cardoso, 2004). The objective of COTRIN is the implementation of the semantic XML-based OTA specifications. Major airlines, hoteliers, car rental companies, leisure suppliers, travel agencies and others will use COTRIN to bring together autonomous and heterogeneous tourism web services, web processes, applications, data, and components residing in distributed environments.

The LA_DMS (Layered Adaptive semantic-based DMS and P2P) project deployed the tourism destination ontology to enable destination management systems (DMS) adaptive to user’s needs concerning information about tourism destinations (Kanellopoulos et al., 2005).

Jakkilinki et al. (2005) provides an overview of development methodology and applications for tourism ontologies. Ontologies are created using ontology development tools, such as Protégé 2000 (Protégé, 2000) that provides to the user: a) construction of a domain ontology, b) customization of data, and c) entry of data. Protégé is a Java-based ontology editor with OWL Plugin: it allows ontology implementation as an applet on the Web. This permits multiple users to share the ontology. The W3C (World Wide Web Consortium) has recently finalized the OWL language (http://www.w3.org/2004/OWL/) as the standard format in which ontologies are represented online. OWL provides greater machine interpretability of web content than that supported by XML, RDF and RDF-Schema (McGuinness & Van Harmelen, 2003). With OWL we can implement a semantic description of a tourism/travel domain by specifying its concepts and the relationships between the concepts.

An example scenario of semantic web

Soon, providers of travel-related services such as accommodation and holiday activities will advertise their services on the semantic web, so that intelligent software agents can find them dynamically. These software
agents could then make suggestions on vacation planning and make travel arrangements in consideration of user preferences. For these agents, the semantic web infrastructure would be based on two core ontologies as illustrated in Figure 1. Both ontologies would be published on fixed URI’s (Universal Resource Indicators) as OWL files. The travel ontology would allow providers to publish metadata about their travel services and contact information. Providers would instantiate the classes from the ontology and publish the resulting individuals as OWL files on their web sites. Then, a semantic web service specialized in vacation planning could send out a crawler agent to collect the available activities. If a user then asks for an exciting adventure destination, the agent could exploit the categorization of the ontology hierarchy to find suitable matches, and call auxiliary web services via the links into the geography ontology. Providers of activities cannot only publish their metadata dynamically, but they can also define their own specializations of the default classes. For example, an ontology module could define HeliBungeeJumping as a subclass of BungeeJumping, and put semantic restrictions on this class to describe its characteristics. Then, if a software agent searches for bungee jumping facilities it would also find the instances of the subtypes, and also learn that jumps from a helicopter are traditionally more expensive than conventional jumps, that they involve aerial sightseeing, etc.

**Figure 1. Ontologies in a scenario**

![Ontologies in a scenario diagram](image-url)
APPLICATIONS OF TOURISM ONTOLOGIES

“Harmonise is an ontology-based mediation and harmonisation tool” (Dell’Erba, 2004), that, in conjunction with other initiatives (EnjoyEurope, Fetish: http://www.fetish.t-6.it/) and with the involvement of international tourism organizations (ETC, IFITT, NTOs, and so on), establishes the bridges between existing and emerging online marketplaces. The Harmonise project allows participating tourism organizations to keep their proprietary data format and use ontology mediation while exchanging information (Missikoff et al., 2003; Dell’Erba, 2004).

The Satine project developed a secure semantics-based interoperability framework for exploiting web service platforms in conjunction with peer-to-peer (P2P) networks in the tourist industry (Dogac et al., 2004). The essence of P2P computing is that nodes in the network directly exploit resources present at other nodes of the network without intervention of any central server. Maedche and Staab (2003) analyzed the advantages of web semantics and P2P computing for service interoperation and discovery in the travel domain. The EU-IST project SWAP (http://swap.semanticweb.org/) demonstrated that the power of P2P computing and the semantic web could actually be combined to share and find “knowledge” easily with low administration efforts. The LA_DMS project provided semantic-based tourism destination information by combining the P2P paradigm with semantic web technologies (Kanellopoulos & Panagopoulos, 2005). Semantic web methodologies and tools for intra-European sustainable tourism were developed in the Hi-Touch project (Hi-Touch, 2003). These tools are used to store and structure knowledge on customers’ expectations and tourism products. The top-level classes of the Hi-Touch ontology are documents, objects and publication. Documents refer to any kind of documentation, advertisement, about a tourism product. Objects refer to tourism offers themselves, while a publication is a document created from the results of a query. Machines and users can process the knowledge on customers’ expectations and tourism products in order to find the best matching between supply and demand. The Hi-Touch platform has already been adopted in several French regions.

Tourism information semantic modeling and querying

Semantic annotation is the process of inserting tags in documents in order to assign semantics to the text. The success of the semantic web in
the tourism industry will depend on the availability of suitable ontologies as well as the proliferation of web pages annotated with metadata conforming to these ontologies. Kiryakov et al. (2004) proposed various promising techniques for semantic annotation, indexing and retrieval of such web pages. However, the presentation of these techniques is out of the scope of this paper. Figure 2 shows a basic architecture of an annotation environment. The document editor/viewer visualizes the documents contents and supports various formats. The metadata creator provides new metadata easily by selecting pieces of text and aligning them with parts of the ontology. The annotation tool GUI also allows the controversial authoring of documents with the aid of the ontology browser. Instances already available may be dragged from a visualization of the content of the inference engine and dropped into the document. A good visualization of the ontology helps to correctly choose the most appropriate class for instances. The inference engine reasons on crawled and newly created instances and on the ontology. It is used to query whether and which instances already exist in the semantic web and it serves the ontology browser, because it allows querying for existing classes, instances and properties. Usually the most inference engines are implemented using the Racer tool (Racer Reasoner, 2004). During the metadata creation, subjects must be aware of which entities already exist in the semantic web.

**Figure 2. Architecture of annotation environment**

![Architecture of annotation environment](image-url)
The semantic modelling of tourism information enables intelligent tourism information systems to provide personalized services. An intelligent tourism information system includes ontology-driven subject domain and repository of tourism information. It is adaptive to user’s needs (e.g. a user requires to be informed about transportation, restaurants, accommodation, services, weather, events, itinerary tips, shopping, nightlife, daily excursion, car rental, sport activities…). Information management tasks are annotated in terms of subject domain concepts which are used as a basis for implementing intelligent system’s adaptive behavior. The system’s adaptive behavior to users’ needs is obtained by attaching semantic metadata to its information modules. For achieving this, tourism concepts ontologies (being used) must be also aligned with the ontologies defining its context and the user’s profile. The system’s adaptability requires the tourism information of the knowledge base to be modeled using multiple descriptions (viz. using various templates associated with the user’s needs). In the LA_DMS project, Kanellopoulos et al. (2005) proposed a layer-based approach for semantic labeling of a tourism destination information. The layers of their semantic labeling reflect a higher level of semantics and constitute sub-models, such as tourism destination model, user’s model (user’s preferences) and machine’s model (e.g. presentation properties). As a result, the LA_DMS model enables DMS to provide personalized information services for tourism destinations.

Semantic querying for tourism information

The need for searching information is one of the fundamental needs of a prospective tourist. Maedche and Staab (2002) presented semantic search scenarios for tourism information. Semantic search enhances current search engines with semantics: It goes beyond superficial keyword matching by adding semantic information, thus allowing easy removal of non-relevant information from the result set. Semantic search can be provided by semantic web tools, such as the Ontobroker system that provides an ontology-based crawling and answering service (Fensel et al., 1999).

Semantic browsers, such as Magpie (Dzbor et al., 2003), use ontologies to identify important concepts in a document and provide access to relevant material. Semantic browsing locates metadata and assembles point-and-click interfaces from a combination of relevant information: It should be able to allow easy navigation through resources, since users with any level of computing knowledge may use it.
SEMANTIC PORTALS AND SEMANTIC WEB SERVICES

Existing tourism portals on the Web have the limitation that they only present accommodations and tourism facilities that are in their databases. Furthermore, these portals rely on existing web technologies that are not able to perform efficient searches—really giving the users what they need. A tourism knowledge (semantic) portal can be seen as a web application providing access to tourism data in a semantically meaningful way, making available a variety of tourism resources for diverse target audiences. Differently from “dumb” web portals, semantic portals are “smarter” and carry out intelligent reasoning behind the scenes. They should offer semantic services including semantic-based browsing, semantic search and smart question answering. Knowledge portals provide views onto tourism information on the Web, thus facilitating their users to find relevant specific information.

The OnTour project (Prantner, 2004) built a semantic portal that searches semantically annotated websites and retrieves efficient and optimal results using semantic web technologies. The KAON portal (http://km.aifb.uni-karlsruhe.de/kaon/Members/tvo/kaon_portal) is a simple tool for generating of ontology-based web portals. To create the portal, the user needs to create an ontology containing the information, which will be presented on the Web. Then, the KAON portal may be used to provide default visualization and navigation through this ontology. There is also the SEAL portal (SEmantic portAL) that exploits semantics for providing and accessing information at a portal as well as constructing and maintaining it (Maedche et al. 2001).

The aim of semantic web services is to describe web services’ capabilities and content in a computer-interpretable language, and improve the quality of existing tasks, including web services discovery, invocation, composition, monitoring, and recovery (Sycara et al., 2003). They have major impact on the tourism industry as they allow the automatic discovery, composition, integration, orchestration, and execution of inter-organization tourism business logic, making the Internet become a global common platform (McIlraith et al., 2001). Tourism semantic web services can constitute: 1) the automated identification of tourism information, 2) the semantic discovery and interoperability of tourism web services, 3) the personalized tourism web services, and 4) the P2P-based semantic web services. Sakkopoulos et al. (2006) proposed techniques to facilitate semantic discovery and interoperability of web services that manage and deliver web media content. In addition Kanellopoulos et al. (2004) proposed a novel
management system of semantically enriched web travel plans. This system evaluates how on-line travel plans are consumed and identifies the individual differences among the users in terms of travel plan content usage.

**E-MARKETS AND INTELLIGENT SOFTWARE AGENTS EXPLOITING SEMANTICS**

In the tourism industry, new offers and requests typically come in by the minute and late vacancies of rooms, flights or lodging easily can be lost. Therefore, there is a need for a fast match between providers and requestors. In e-markets that exploit semantic descriptions, semantic-based matching of products and requirements is made fast between tourism providers and requesters, while a large volume of transactions is executed.

Sycara et al. (1999) described a comprehensive software agent framework that allows the set up of semantic-based e-markets. In semantic-based e-markets, intelligent software agents can exploit semantics on the Web. Actually, the semantic web can utilize a variety of traveler, hotel, museum and other software agents to enhance the tourism marketing and management reservation processes (Hendler, 2001). For example, a hotel software agent operating on the semantic web might undertake many of the routine administrative tasks that currently consume large amounts of a hotel manager’s time. Also, traveler software agents can assist travelers in finding sources of tourism products and services and in documenting and archiving them. An additional capacity of the semantic web is realized, when software agents extract information from one application and subsequently utilize the data as input for further applications. In this way, software agents can create greater capacity for large scale automated collection, processing and selective dissemination of tourism data.

**Dynamic packaging systems**

The Web has permanently changed the manner vacation packages can be created. Consumers can now acquire packages from a diversity of websites including online agencies and airlines. In the travel industry, one of the fastest-growing categories is the creation of dynamic vacation travel packages. The objective of dynamic packaging is to pack all the components chosen by a traveller to create one reservation. Regardless of
where the inventory originates, the package that is created is handled seamlessly as one transaction, and requires only one payment from the consumer. Dynamic packaging systems create customized tourism packages for the consumers. A dynamic packaging application allows consumers or travel agents to bundle trip components. The range of products and services to be bundled is too large: guider tour, entertainment, event/festival, shopping, activity, accommodation, transportation, food and beverage etc. Dynamic packages can be created and booked effortlessly with private and published air, car hire, hotels, attractions and insurance rates. It is remarkable that dynamic packaging platforms can be deployed, if we use only semantic web technologies (Cardoso, 2005).

**Semantic mining**

Semantic data mining allows precise targeting, personalization of tourism products, and measurability; viz. tools for effective tourism marketing strategies. For example, semantic mining can be very useful for the tourism destinations management or the travel plans management (Kanellopoulos et al., 2004). Semantic mining process can be applied to record and analyze users’ preferences concerning in specific elements of a tourism information module. Intelligent tourism information systems can generate users’ profiles by recording users’ preferences. A user profile is used for expressing the characteristics and features of a person. It consists of a static part (e.g. demographic info such as name, sex, age, country of origin etc) and a dynamic part (interests, filters, traces). Filters describe the mechanism for expressing user’s interests. For example, a filter expresses the fact that a user is interested in museums. Traces describe the interactions of users with the tourism information system and the mechanism for recording these actions. Future Internet marketing policies will be based on the usage rate of tourism semantically content items (in websites) and will be related to the individual differences among users regarding content items consumption. As the main dependent variable can be used the notion of “content item view” $C_{ij} \in \{0,1\}$, which indicates whether user $i$ ($i=1\ldots n$ users) clicked on a link of a content item $j$ ($j=1\ldots m$ items) and accessed it.
SEMANTIC WEB SOCIOLOGICAL IMPLICATIONS FOR DESTINATION MANAGERS

With the spread of the first computers we believed that as machines replace humans, we will interact with them more that with each other, making the world less of a social space. Paradoxically, it seems that nothing could be less true: we shaped our information systems to our form and move much of our social life in the electronic domain. In the area of social software, we find techniques for extracting, representing and aggregating social knowledge.

In fact, destination management organizations (DMOs) or destination managers constitute a social network as they are connected by a set of social relationships, such as co-working and information exchange. Using social network analysis (Wasserman et al., 1994), patterns that represent tourism destination networks and associations between destination managers can be constructed automatically. Such an analysis could yield the main groups of destination managers and identify the subgroups, the key individuals (centrality) and links between groups. Network analysis can benefit destination managers’ communities by identifying the network effects on performance and helping to devise strategies for the individual or for the community accordingly. In terms of social network analysis, the use of electronic data provides a unique opportunity to observe the dynamics of destination managers’ community development.

In the semantic web framework, the “Friend-of-A-Friend” project (FOAF: http://www.foaf-project.org) can represent social networks and information about people (user profiles) in a machine processable way. The FOAF project is highlighted by the following features: a) publishing personal profile with better visibility; b) enforcing unique person identity reference on the Web and thus supporting the merge of partial data from different sources; and c) representing and facilitating large scale social networks on the Web.

For the extraction and analysis of online social (tourism destination) networks we can use the Flink system (Mika, 2005). Flink can employ semantic web technology for reasoning with “personal” destination information extracted from a number of electronic information sources including web pages, emails, etc. The acquired knowledge can be used for the purposes of social network analysis and for generating a web-based presentation of the tourism destination community. In addition, the Flink exploits FOAF documents for the purposes of social intelligence. By social intelligence, we mean the semantics-based integration and analysis of social knowledge extracted from electronic sources under diverse
ownership or control. Conclusively, Flink is interesting to all destination managers, who are planning to develop systems using semantic web technology for similar or different purposes.

In the near future, two great challenges are going to emerge in the tourism industry: 1) creating a social ontology for destination managers that would allow classifying complex, social relationships along several dimensions; 2) finding patterns for identifying these relationships using electronic data. As destination managers’ lives become even more accurately traceable through ubiquitous computers, the opportunities for social science based on electronic data will only become more prominent.

CONCLUSION

Currently, the tourism industry is facing rapid changes with the advent of the semantic web technologies. For example, a semantic web application allows consumers or travel agents to create, manage and update itineraries. Moreover, it permits the customer to specify a set of preferences for a vacation and query a set of information sources to find components such as air fares, car rental, and leisure activities in real-time. Intelligent tourism information systems offer full integration, flexibility, specialization and personalization.

Full Integration: Intelligent tourism information systems can integrate the management and marketing of the various local tourism products and services (Bussler, 2003). They can facilitate interconnectivity of Small and Medium Tourism Enterprises (SMTEs) via full integration in order to increase margins on the products sold. Tools for sales assistance, such as ‘intelligent’ software agents, can provide various products and services into an integrated tourism package, which is personalized to tourist’s needs.

Flexibility: Intelligent tourism information systems can combine the individual tourism products and services. They are platform independent and can change their data without affecting the data representation.

Specialization and personalization: Precise targeting, personalization, privacy and measurability can be achieved through web direct marketing that is interactive, immediate, and accurately timed. Through web direct marketing, tourism products and services can be personalized to the user’s needs (Murphy, 2003). Finally, the utilization of intelligent tourism information systems offers better information management and achieves automatic intra (or inter)-organizational communication of a higher quality.
The semantic web forms a platform for search engines, information brokers and ultimately the ‘intelligent’ software agents. It propagates interoperability, reusability and shareability, all grounded over an extensive expression of semantics with a standardized communication among intelligent tourism information systems. There is now the need for developing an infrastructure to manage the online tourism information and deliver to consumers what they want. New superior consumer services can be deployed such as tourism market overview and price comparison. Ontologies will play an important role as they promise a shared and common understanding of tourism and travel concepts that reaches across people and application systems. Semantic-based tourism information systems will revolutionize the tourism industry. Despite, the methodology of applying the semantic web in intelligent tourism information systems needs to mature and methods for achieving scalability and robustness need to be developed.

REFERENCES


Murphy, H.C. (2003). An Investigation into how data collected by destination websites are utilised as a direct marketing tool. In A. Frew (Ed.) Information and Communication Technologies in Tourism, Helsinki: Springer, pp.316-325.


**SUBMITTED: OCTOBER 2005**
**REVISION SUBMITTED: MARCH 2006**
**ACCEPTED: JUNE 2006**
**REFEREED ANONYMOUSLY**

**Dimitris Kanellopoulos** (dkanellop@teipat.gr) is a Lecturer at the Technological Education Institute of Patras, Department of Tourism Management, Meg. Alexandrou 1, Patras GR-26334, Greece.