Knowledge discovery and computerized reasoning to assist tourist destination marketing

George Stalidis [a], Dimitrios Karapistolis [a]

[a] *Department of Business Administration, Alexander Technological Educational Institute, Thessaloniki, Greece*

Abstract: In this paper we present a knowledge-based approach where data analysis and reasoning technologies are consolidated in order to provide solutions to common marketing planning problems. The aim was to support marketers in the tourist domain with intelligent electronic tools, which are designed to extract knowledge from survey data, maintain it in a suitable Knowledge Base and offer it for problem solving in a computerized environment to users who are not experts in analysis. The focus was to capture the image of Thessaloniki as tourist destination, to study the characteristics and needs of its visitors and to suggest important factors for improving the promotion of the city to individual visitor groups. The data sources were questionnaire-based primary surveys on the destination image and the expectations of visitors regarding their hotel. Multidimensional Multi-factorial Data Analysis was used as a knowledge extraction method to discover factors, associations and other hidden patterns that explain the behavior of tourists, as well as clusters that can be studied individually. An ontology-based Knowledge Modeling process was then developed to express the results of data analysis as machine understandable rules, so that an inference engine can respond to high-level queries. Special types of visitors were identified, such as visitors for nightlife and visitors for culture, and rules were formulated to predict the requirements of such types, in relation to parameters such as age and country. A query mechanism running on rule-based reasoning was successful in providing suggestions to typical marketing problems.

Keywords: Knowledge extraction from data, data analysis, destination marketing, knowledge modeling, computerized reasoning.

1. Introduction

Data analysis and Decision Support systems are promising tools for marketing planning, enabling marketers to make informed decisions, such as the selection of target market segments, product positioning or optimal configuration of campaigns. A challenging application area is the one of tourism, which evolves rapidly and addresses a complex market. Since the access to high quality market information and the ability to take successful intelligent decisions may be of decisive importance for the competitiveness of a destination, it is evident that information technologies can become valuable tools, not only for information search and data processing but also on the level of sophisticated analysis and effective usage of specialized knowledge. Since more than two decades ago, it has been supported that successful

marketing in tourism depends on the extent to which more specialized consumer demands or lifestyles can be identified, as opposed to massive generic approaches [1,2]. The positioning strategy of a tourism product can thus be devised following the measurement of the customer's image of the tourism product [3] and his satisfaction from product attributes, in correlation with his needs and desires [4]. Managing tourist destinations is a challenging field, where additional factors are involved, such as culture, branding and communication [5].

Measurements of a destination's or product's image as well as the identification of the customer's needs are typically performed through primary surveys and statistical analyses. Considering that these primary data sources are often limited or costly, it would be of great value to make the most out of available survey data by discovering the knowledge hidden in them, by sharing and reusing as much as possible the survey results, as well as consolidating their findings with other surveys and secondary data for solving complex problems. Moreover, it would be desirable for a marketer to be able to access the knowledge derived from existing surveys by posing high level queries to an intelligent mechanism instead of dealing with datasets, statistical tables or reports. The challenge in achieving these goals is twofold [6]. Firstly, specialized knowledge extraction methods [7] need to be employed to discern qualitative patterns and hidden dependencies, as well as to study phenomena evolving over time. The second challenge is to model the analysis results so that these can be accessed by an intelligent decision support engine and applied to computerized problem solving [8,9].

One form of electronic knowledge management is the use of information systems to classify an organization's knowledge and to manage global or individual knowledge repositories. Additionally, it is possible to capture ideas and communication in discussion forums, chat sessions, live meetings, and interactive polls. The purpose of systems in this category is to record and organize knowledge, so that this knowledge is not lost, it is accessible when needed and searchable through clever engines, either locally or over a network. An important feature in this case is that the information system is concerned with the management of knowledge (e.g. indexing and meta-tagging, archiving, etc.) but not with the content itself. The actual knowledge can only be appraised by the person who is responsible for performing a task e.g. a marketing executive by reading and applying it.

In this work we focus on a more sophisticated approach of electronic knowledge management, usually referred to as Knowledge Engineering (KE), which was defined in 1983 by Edward Feigenbaum and Pamela McCorduck [10] as an engineering discipline that involves integrating knowledge into computer systems in order to solve complex problems normally requiring a high level of human expertise. The ability to infer and suggest solutions by utilizing knowledge is passed from the human analyst to an inference engine. Although this approach has the limitation that a computer system would never fully replace a human expert, there are several strong points in KE that make the effort worthwhile.

Firstly, knowledge engineering [9] facilitates the storage and consolidation of large amounts of knowledge (e.g. thousands of rules and item classifications) which can be applied simultaneously by an intelligent system to solve complex problems, while on the contrary it would be impossible for an expert to consider as a whole. Secondly, continuous updates of electronic knowledge are possible by introducing new findings and by correcting or discarding obsolete ones, taking advantage of algorithmic methods for automatic checking against inconsistencies. Reasoning engines can be used to produce inferred knowledge from the declared one by combining axioms, background domain knowledge and problem-specific knowledge,

even considering the context within which part of the knowledge is applicable. In this way, the knowledge capital becomes coherent, maintainable and reusable. Thirdly, and maybe most importantly, the knowledge accumulated in a Knowledge Base (KB) can be used through intelligent decision support tools – incorporated into a Knowledge-Based System (KBS) - to provide solutions to complex problems, without requiring expertise or deep comprehension by the user. By adopting the appropriate formalism, knowledge can be exchanged between systems or accessed over the Internet.

Strategic planning tools for improving the competitiveness of tourism in selected areas have been reported in the literature [11], based mainly on information management and less on sophisticated analysis and knowledge extraction. Knowledge-based decision support systems applied to tourism marketing have also been reported [12]. In the current paper we addressed these goals by developing a knowledge-based Tourist Marketing Decision Support System (TDSS), focusing on the design of a knowledge model suitable for tourist destination marketing and proposing a methodology for managing knowledge derived from multiple questionnaire-based primary surveys. Multidimensional factor and clustering analysis methods [13, 14] were used as a powerful knowledge extraction method and original results were derived from a recent marketing survey on the image of Thessaloniki as a tourist destination. A Knowledge Base was then developed employing ontologies and rule-based knowledge engineering technologies [9], implemented using the OWL semantic web ontology language [15]. The Knowledge Base was populated with both background knowledge on tourist destination marketing and findings from the analysis of questionnaire-based surveys. Finally, a reasoner [16] was applied to infer logical consequences from the asserted facts (axioms and inference rules) and a query mechanism [17] was used to provide suggestions and predictions based on the accumulated knowledge.

In previous work, the authors proposed an initial version of the Knowledge Model and populated the KB with preliminary results from the surveys on the image of Thessaloniki and on hotel satisfaction [18,19]. They also applied multidimensional data analysis in parallel with data mining as alternative knowledge extraction methods [20]. In the current work, the final results of the full survey are used to introduce richer content in the KB and the problem solving capabilities of our approach are illustrated by applying a query mechanism to obtain computed suggestions to simple problems. In the following Section 2, we present two primary surveys that were conducted on the Tourism of Northern Greece and the results derived by employing multidimensional factor analysis. In Section 3, the proposed KE approach is presented, together with the developed model and the content creation process, while the inference mechanism and its application to the marketing decision support are discussed in Section 4. The conclusions and plans for future work are given in Section 5.

2. Field surveys and data analysis

2.1. Analysis of the image of Thessaloniki and the expectations of its visitors

The wider purpose to be served by the presented knowledge discovery and computerized reasoning approach is the capturing of trends, facts, customer needs and perceptions in the field of tourism, so that these can be employed by a Destination Management Organization (DMO) to enhance marketing planning for the benefit of a specific destination. In order to precise the marketing problem on which the research work was focused, the authors collaborated with the DMO of

Thessaloniki (Organization for the Touristic Development and Marketing of Prefecture of Thessaloniki). The specific organization was involved as an end user of the developments i.e. the one who will utilize the KB and query mechanism, for the purpose of deciding more efficient marketing actions. The application scenario of the envisioned system evolved around three points: (a) analysis of tourist product (image, resources, infrastructure, etc.), (b) analysis of market segments based on their needs, return prospects and accessibility/susceptibility, and (c) support in selecting promising target segments and prioritizing actions to match the selected segments. Decision support was focused on:

- Comparative evaluation of market segments.
- Identification of critical competitive factors of the destination in relation to specific target segment and to discovered or predicted trends.
- Suggestions for optimized strategic planning.

Although the input data sources in the proposed framework can be both primary and secondary data (e.g. data provided by statistical agencies of collaborating agencies), the focus of the current work was on the extraction of knowledge from primary survey data, by applying multidimensional factor analysis methods. To this end, a primary survey was designed and launched in the city of Thessaloniki, specifically for the purposes of the current project. The aim of the survey was to collect information on the perceived image of the city of Thessaloniki, as expressed by different types of visitors and the analysis of the factors contributing to the formation of this image. The survey was addressed to tourists or other foreign visitors of the city, who had already spent at least one night. The respondents were travelling by themselves or in touristic groups and in many cases they were visiting Thessaloniki just for one or two days, on their way to their main vacations in nearby locations, usually beach areas in Halkidiki or Pieria. The instrument was a 3-page structured questionnaire containing 43 questions, organized in 8 sections, including questions regarding the visitor's familiarity with the destination, his general satisfaction and future stance, the reasons for choosing this destination and factors influencing this decision, a number of attributes related to the perceived image of the city and the country, as well as personal/demographic information. The sample was around 2000 tourists, reached during the period from May to October 2013.

In order to uncover the relations among different aspects of the visitors' image for a destination, we applied multi-dimensional factor and clustering methods. In specific, a combination of Multiple Correspondence Analysis (MCA) [13] and Hierarchical Cluster Analysis (CHA) based on Benzecri's chi-square distance and Ward's linkage criterion [13, 21] were used in a multistep analysis procedure. The specific analysis methods have been preferred against quantitative methods because of their strengths (a) to deal with datasets that contain categorical variables and hierarchical ranking variables, without the need to quantify them by means of scales/scores (b) to detect complex relations among a large number of variables without a-priori assumptions on the underlying models, even if the relations are not linear and (c) to produce results in qualitative form, allowing graphical exploration and formation of patterns that involve classes, properties and associations [21]. The data analysis process applied to the questionnaire-based survey data was performed using the MAD analysis software [22].

The first step of the process was to analyze each section of the questionnaire, corresponding to an individual aspect of the survey. The clustering method divided the respondents into homogeneous groups and the factor analysis depicted associations among properties, thus revealing the structure of the responses within each section. The resulting groups of associated properties lead to the definition of

classes which can be treated as categories of new qualitative variables. By combining the two methods, it was possible to associate the groups of respondents with properties and/or classes.

Table 1. The frequencies of responses for image attributes.

Image Attributes	-2	-1	0	1	2	No response
Δ1 Cleanness	100	340	578	597	285	47
Δ2 Natural beauties	24	101	323	809	635	55
Δ3 Prices	65	167	582	755	294	83
Δ4 Sightseeing	30	86	311	727	704	89
Δ5 Greek cuisine	43	103	251	659	811	80
Δ6 Nightlife	73	98	430	608	532	205
Δ7 Architectural style	54	138	365	753	562	75
Δ8 Security	35	135	451	753	467	106
Δ9 Friendliness of local people	36	103	275	669	797	67

In Table 1, the responses received for the 9 items on the image of Thessaloniki are shown. The most negative evaluations were for Cleanness (52,3% negative or neutral), while the most positive ones were for Greek cuisine and friendliness of local people. By applying MCA on the generalized contingency table, it was found on the factorial plane 1X2 (97% of inertia) that there were 5 distinctive standpoints regarding the image of Thessaloniki. As shown in Figure 1, the first factor differentiated positive responses (on the right side) from negative responses (on the left) while the second factor differentiated the negative responses qualitatively. The parabolic curve shows that the responses are consistent with an overall escalation from negative to positive [14]. The most negative standpoint was class A: negative image for sightseeing, Greek cuisine and friendliness of local people, and next were Class B: negative image for prices and security and Class C: negative for natural beauties and city style. Class D corresponded to neutral image for all aspects but negative for cleanness and Class E to positive image for all aspects. The fact that the negative image for cleanness was associated with neutral responses on other items showed that this particular problem was serious, since it was brought up as negative even by visitors who were in general positioned as neutral. Conversely, the negative image for sightseeing, Greek cuisine and friendliness of local people was at the negative end of the scale, which means that it was an extreme atypical standpoint, far from the center of gravity. The cluster analysis resulted in 5 groups of visitors (noted as K1 to K5). The largest in size group (K5) represented 32,3% of the sample and was associated to neutral responses to all items, while group K4 (31%) was associated with positive responses to all items.
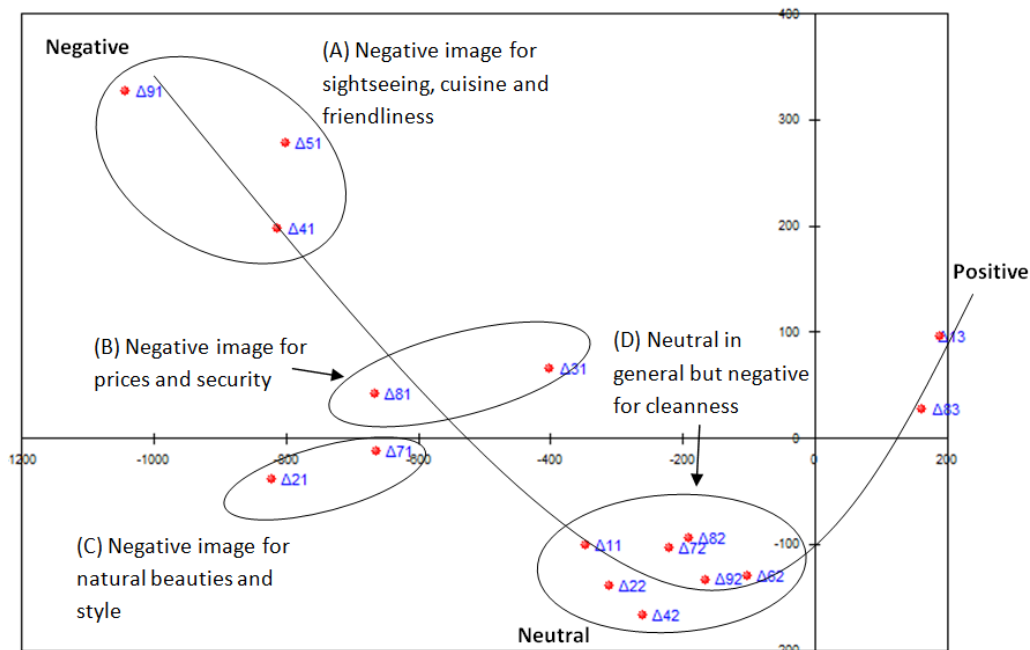
Figure 1. The factorial plane 1X2 of the multiple correspondence analysis of the image of Thessaloniki.

In a similar fashion, the analysis of the visitors' image for the country showed that a group of respondents representing 37,6% of the sample was associated to the most positive responses to all aspects, a group of 9,4% was neutral to most items and negative regarding security in the country, a group of 45% were neutral in everything but diverse regarding historical monuments.

As regards the section on priorities for selecting their destination, four different classes were identified, specifically: (1) the ones who were attracted by the fame and history of the area, they were mainly interested in visiting museums and considered the infrastructures as important factor (26,9% respondents classified in this class), (2) those who were attracted by night-life and life-style, Greek cuisine and were influenced by friends (8,73%), (3) the ones who were attracted by the natural beauty of the area, the climate, natural environment and opportunities for excursions (54,3%) and (4) the summer tourists who were mainly interested in the beaches/swimming (9,9%).

After completing the above analysis process for each section separately, the next step was to apply MCA on the full set of class membership variables in order to observe an overall picture of the dependencies among all structural variables. Personal information/demographic variables were also added as supplementary data in order to associate these characteristics with visitor classes. Results showed that e.g. French visitors were influenced mostly by the opinions of their friends and the climate, were interested in night-life and food and as second priority had the visits to museums, natural beauties and lifestyle. On the contrary, Russian visitors were mainly influenced by the reputation of the destination and the natural beauties and were interested in excursions, contact with nature and the history of the area, while they considered infrastructures and transports as important factors. Both French and Russian visitors, together with those from Italy, Spain and Portugal, had very positive image regarding the friendliness of local people, the sights, the food and the style of the city, while they were neutral regarding cleanness, prices and security. By focusing on specific subgroups and analyzing the relations among selected variables (e.g. duration of visit, income, satisfaction, etc.), more specialized findings were

derived regarding the preferences, loyalty and factors of satisfaction for the visitors of Thessaloniki.

2.2. Analysis of the expectations and the satisfaction of visitors from their hotel

As additional knowledge source, we used a survey on service quality of hotels, addressed to tourists of Northern Greece. The analysis of this survey has been performed in previous work by the first author of this article [20] using both multidimensional factor analysis and data mining techniques. In the present work, the analysis results are briefly presented and it is shown how they were introduced in the TDSS as knowledge content.

The questionnaire was designed to collect information about the expectations and satisfaction of the visitors regarding their hotel, including also the demographic characteristics of the visitors and information about the type of their trip (e.g. duration, cost, etc.). In total, the questionnaire consisted of 21 closed-type questions, including two multi-item ones where the respondents indicated on a five-point scale the degree of their expectation and their satisfaction for each one of 33 attributes. The main purpose was to evaluate the perceived quality of the hotels in the area and to study the factors explaining satisfaction. The aims of the analysis were to identify representative visitor classes in terms of their demographic characteristics and purpose of their visit and to associate these classes with priority expectation attributes. In this way, it was possible to extract rules that predict visitor satisfaction, options and requirements, so that the marketer is able to decide which hotel characteristics deserve more emphasis, given their target customers or, vice-versa, which customers are worth aiming at, given the characteristics of a hotel.

The dataset included a sample of 400 visitors. From those, 41% were men and 59% women, 33% were aged from 26 to 35 years old, 55% were from 36 to 55 and 19% were older than 55, while most of the respondents (67%) had completed higher education. Most of the visitors stated that the purpose of their visit was vacation and just 8% mentioned professional reasons.
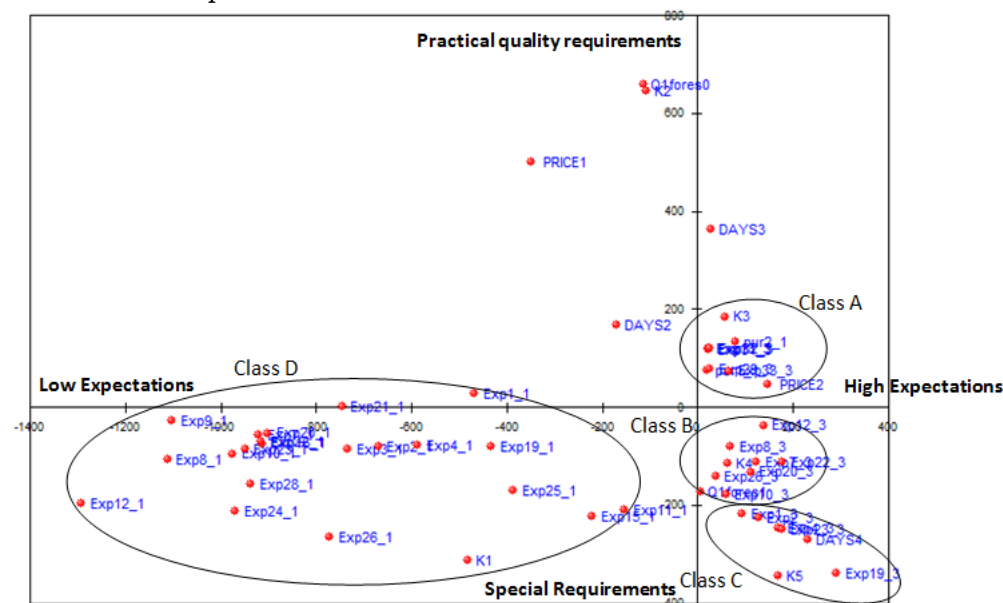

Figure 2. The factors of visitors' expectations from their hotel and the four identified classes.

In order to identify the dominating trends regarding expectations, MCA was applied on the variables related to the characteristics of the visit (e.g. reason for

visit, duration, cost category) and specific expectations from the hotel. On the factorial plane 1X2, as shown in Figure 2, the 1st axis (29,7% of inertia) represented the contrast between low expectations / low cost and high expectations, while the 2nd factorial axis (11,8% of inertia) differentiated expectations for basic quality characteristics from more specialized demands (e.g. sports facilities and spa). On the factorial plane, four groups of properties were visible, corresponding to the following classes:

Class A. In this class, the purpose of the visit was vacation or entertainment, the duration of the trip was around 2 weeks and the cost of the room that the visitor was willing to spend per night was 50 to 100 €. Additionally, there were high expectations for security, cleanness, materials and consumables (i.e. shampoos, towels, etc.), hairdryer and restaurant in the hotel. This class fits the expectations of a visitor who comes for relaxing vacation and is mainly concerned about quality in practical issues. It was named as Standard quality vacation.

Class B. It includes the responses that the visitor has been in this destination once or twice before and his main expectations were swimming pool, entertainment activities, comfortable lobby and the hotel to belong to a group of hotels. This class was named Relaxed vacation.

Class C. It includes the responses that the visit duration was 3 weeks or more and there were requirements for spa and wellness services, facilities for persons with special needs, sports facilities and special diet menu. This class was named Activity and wellness and fits the profile of a visitor with specialized needs for high quality vacation.

Class D. It includes low expectations overall and willingness for the lowest hotel price. It is the Low budget class.

Following the above analysis, CHA was applied to associate visitor groups with the above classes. It came out that the largest group K1 (N=182, 43,4% of sample) was associated to Class D - Low budget, while Class A - Standard quality vacation corresponded to a group with size 8,1% of the sample. In order to find the demographic characteristics of the identified groups, MCA was applied on the cross-tabulation of the group membership variable with the demographic variables (e.g. age, income, etc.). The results included that the First time visitor was associated to ages 26 to 35 and the Standard quality vacation and Relaxation classes had similar demographic profile which was characterized by age 56-65 and profession retired. The Activities and wellness was characterized by middle ages 46-55, working as freelancer with medium-low income. The next analysis step was focused on satisfaction attributes. It was found that First time visit was associated to low satisfaction from room service, cleanness, quality of food and security. On the other hand, the statement that the visitor would definitely recommend the hotel was associated to family-type hotel and high satisfaction from cleanness, special diet menu, quality of food, spa and relaxation, swimming pool for children and athletic facilities.

3. Knowledge Engineering

3.1. Knowledge-based framework and technologies

The design and development tasks of a Knowledge-Based System and the stages involved in acquiring and using knowledge in computerized form are several [13]. The most important aspects in KE are Knowledge Extraction, Knowledge Modeling and Inference, corresponding respectively to the following three most crucial tasks: to produce (or discover) knowledge from data (e.g. through statistical analysis or

data mining), to express the knowledge in a formal, standardized and machine-understandable form, and to use the accumulated knowledge for problem solving. There exists however an overall consensus that the process of building a KBS may be seen as a modeling activity [9] i.e. building a computerized model with the aim of realizing problem-solving capabilities comparable to a domain expert. Available modeling approaches include ontologies, statistical models, neural networks, rule-based models, case-based reasoning models, each one offering different level of expressiveness and suitability to different kinds of problems. Among other modeling frameworks that became prominent within Knowledge engineering, such as Common KADS [23] and Model-based and Incremental Knowledge Engineering (MIKE), in this work we adopted the ontology-based framework Protégé [24]. An Ontology provides the basis for building a model of a domain, defining the terms inside the domain and the relationships between them [25] and is thus a powerful and widely adopted tool, not only for developing models but also for communicating structured knowledge e.g. through semantic web or otherwise. There are different types of ontologies including Domain ontologies, Generic ontologies, application ontologies and Representational ontologies. In a recent review of existing ontologies in the tourism sector [26] it was found that there are considerable number of efforts, such as the QUALL-ME [27] and the DERI e-tourism [28] ontologies.

As the basic platform for the development of the Knowledge Model and the implementation of the Knowledge Base and the decision support mechanism, we used the Protégé OWL 4.2 [24]. The specific version of this platform is based on the Web Ontology Language [15], which supports semantic definitions and inference mechanisms that allow not only the expression of known facts but also the derivation of their logical consequences to produce knowledge that was not explicitly introduced. OWL has been chosen for its ability to incorporate logic and its suitability for exchanging knowledge through the semantic web. OWL was available in three sublanguages of increasing expressiveness (OWL Lite, OWL Description Logic and OWL Full), from which the OWL DL was used, as more suitable to logical inference problems. The SWRL (Semantic Web Rule Language) [29] was also used to compile rules and the SPARQL language [30] was used to query the Knowledge Base. Both these languages are supported by Protégé and work on top of the OWL ontology.

3.2. The Knowledge Model for Destination Marketing

The starting point in the model design process was that the model should be appropriate to the available knowledge sources, in our case to be able to express the results of multidimensional data analysis applied on a questionnaire-based survey. Secondly, it should be suitable to the intended use of the stored knowledge, which in our case was to provide suggestions to marketers for improved destination marketing decisions. Finally, the constructed model should be expandable, maintainable and generic enough to capture relative knowledge from additional possible sources, such as secondary data or even theoretical knowledge. Considering that the results of the data analysis used are qualitative i.e. in the form of associations, classes, clusters of individuals and in general logical patterns, the most suitable model to be adopted as a basis was the ontology. The ontology [25] provided a description of the domain of interest, which included definitions of concepts, objects or individuals, classes (i.e. types of objects), relations among them and properties. This provided a standardized semantic base for problem solving, which allowed the exchange of knowledge between systems, the integration of heterogeneous data sources and, more importantly, a formal basis which is the

prerequisite for formal rule statement creation and inferential analysis. In the context of our TDSS, an ontology-based model was developed to cover the knowledge representation needs at the following levels (Figure 3):

(a) To provide a terminology regarding the concepts found in the addressed problem (e.g. visitor, destination, trip, hotel etc.) and their attributes (e.g. a visitor has as attributes his age, country of origin, education, etc.). This terminology provides a standardized vocabulary and is organized in a hierarchical structure that includes classes (i.e. types of objects), subclasses (i.e. more specific types which inherit the properties of their superclasses) and individuals (which may belong to certain classes and may have certain properties). For example, Hotel and Camping are both specific types of Accommodation, while the individuals RoomService, Internet and Swimming pool belong to the class HotelFeature. Part of the ontology contains basic definitions of the domain that can be considered as static and another part is a problem-specific one that expands the above basic domain ontology to support the special terminology needs of individual knowledge sources. This component includes definitions of previously unknown terms introduced in a particular survey and definitions of special classes to express some of the findings of the analysis, such as visitor profiles (e.g. a newly identified category of visitor who has negative image regarding prices and quality of services but positive regarding sights and culture)

(b) To enable the expression of relations between objects and the assignment of specific values to object attributes. For example, the individuals Italy, Germany and Russia belong to the class Country and Smith belongs to the class Visitor. The property isFrom has as domain the Visitor class and as range the Country class, which means that it can link any visitor with any country. The property hasAge can be defined to assign an integer value to any Visitor. Using these definitions we can express that Mr Smith is a 30 year old Italian visitor by asserting in the ontology the properties: Smith isA Visitor, Smith isFrom Italy, Smith hasAge 30. The declaration of such relations is formulated on the basis of the vocabulary mentioned above and may express certain knowledge acquired from the analysis, such as associations between profiles found on factorial axes.

RULES
- Associations and classifications derived from Data Analysis.
- Logical relations, conditions and predictions among individuals and properties.

Model type: Rule-based
Implementation: SWRL

Problem-specific relations
- Properties to express relations and attributes.
- Descriptions of composite classes

Model type: Ontology
Implementation: OWL-DL

Tourist domain terminology

- Established general definitions in the domain of tourism

- Additional terminology to match the needs of specific surveys

- Composite classes derived from data analysis
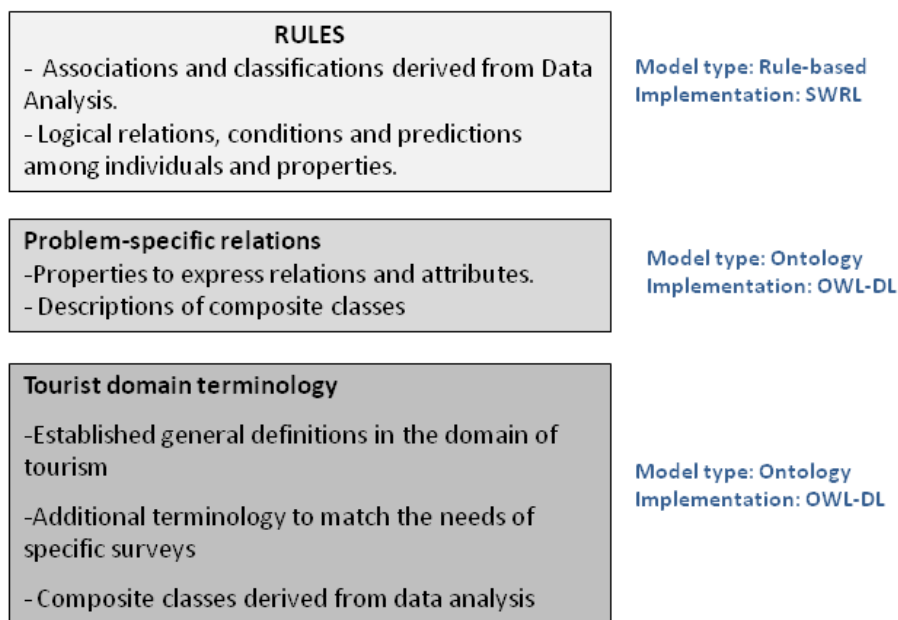
Model type: Ontology
Implementation: OWL-DL

Figure 3. The structure of the Knowledge Model.

(c) To provide a formalism to express more complex associations among concepts, their instances and their properties in the form of rules. This would allow us to incorporate logic and operations, in order to formulate knowledge in the form of predictions or suggestions estimated given a set of conditions and input variables. Considering that the complexity of the knowledge to be engineered is usually higher than class hierarchies and relations between objects, additional expressiveness is offered by rule-based models. A rule-based knowledge framework consists of production rules [31], which are generally expressed in the form:

$$C1 \text{ AND } C2 \text{ AND } \ldots \text{ AND } Cn \rightarrow E(1)$$

where C1, C2, . . . , Cn constitute the conditions of the rule, and E is the consequent, which can be a prediction or suggestion. In our case, rules of this form mainly result from the data analysis and correspond to conditional associations between classes or individuals (e.g. if visitor is young and his purpose of visit is vacation, a first priority decision factor is nightlife) or to classifications (e.g. to characterize a trip as a low-budget trip).

3.3. Knowledge Elicitation and Content Creation

In this section, the developed model and the process of populating it with new inputs are outlined. Figure 4 illustrates the ontology class tree which reflects the consolidated model of both marketing surveys, on the destination's image and on the expectations and satisfaction from hotels. The development of the model and the insertion of analysis findings in the Knowledge Base was a multistep dynamic process which involved interpretation of analysis results, elicitation as well as decisions in the way of representation. The process involved the three levels mentioned in section 3.2.
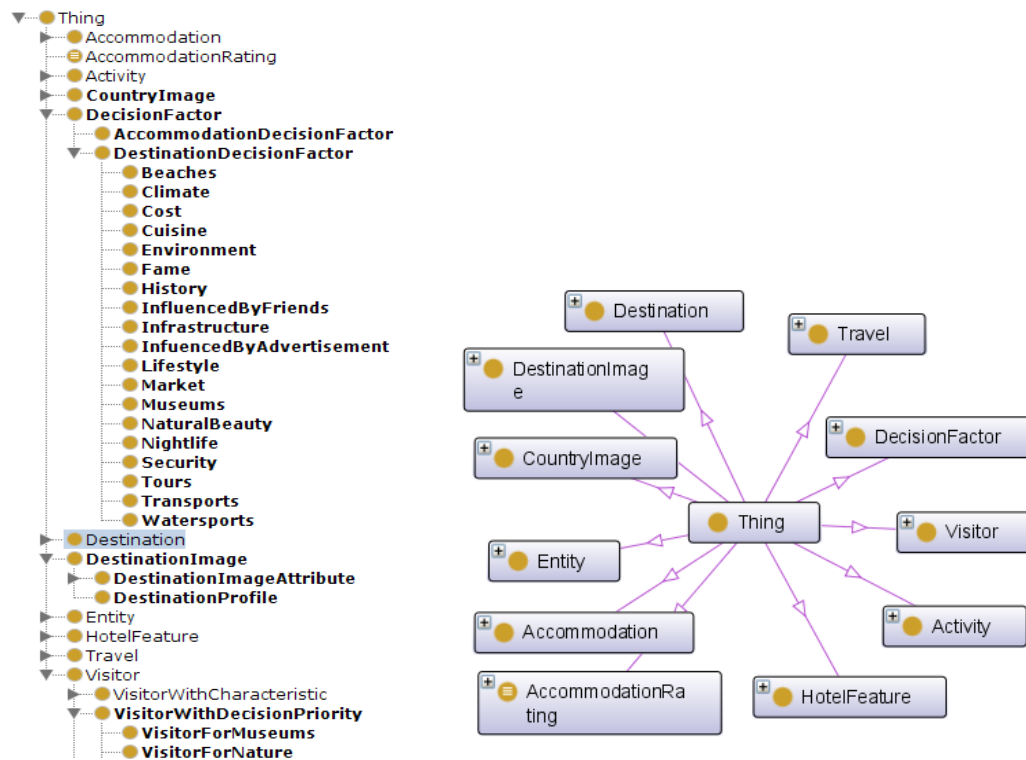

Figure 4. Extract of the ontology class tree (left) and graph of the main classes (right)

a) The basic concepts that were represented as OWL classes were: Visitor, Destination, Accommodation, Travel, *DestinationImage, CountryImage, DecisionFactor* and *Activity*, while other important classes, necessary for representing variables and attributes, were *VisitorFeature, VisitorSatisfaction, HotelFeature* and *TravelFeature. DecisionFactor* includes the subclass *DestinationDecisionFactor*, which includes the subclasses *Beaches, Climate, Cost*, etc. to cover all the available items of the corresponding questions. Each of these subclasses contains as individuals the available answers/categories e.g. *FirstPriorityBeaches, SecondPriorityBeaches*. In addition to this basic level of terminology definitions, an essential element in our modeling approach was to introduce the groups or factor analysis classes that resulted from the Data Analysis stage, in order to be able to describe them in the Knowledge Base and to assign properties to them. For example, *VisitorForNature* was defined as a type of visitor who selected this destination giving first priority to *NaturalBeauty, Climate* and *Tours*. Similarly, classes of visitors were defined according to their image, to their decision priorities, to their expectations from their hotel, and whatever structure has been found by the analysis in the primary data. This type of OWL classes were named "composite" classes, they could only be introduced in the ontology after they were discovered in the data and it is expected that new such composite classes will be dynamically added as the model evolves. At the current stage, we introduced 4 composite classes to express the visitor's destination decision priorities, 6 for the visitors' image, 5 for the visitor's general profile and 4 for the expectations from his hotel.

b) OWL properties were defined to represent relations, such as the property hasRequirement that links a *Visitor* with an *Activity* or *HotelFeature* and the property hasAge that links a Visitor with an age category. Additionally, OWL class expressions were used to describe composite classes as OWL equivalent classes. For example, the class *VisitorForLifestyle* was described as

   VisitorWithDecisionPriority
   and (hasDecisionFactor value *FirstPriorityFood*)
   and (hasDecisionFactor value *FirstPriorityLifestyle*)
   and (hasDecisionFactor value *FirstPriorityNightlife*)

With the above definition, the inference engine understands that any visitor who belongs to *VisitorForLifestyle*, is automatically known to consider Food, Lifestyle and NightLife as first priority decision factor for choosing his destination.

c) The next level was to formulate more complex relations as rules of the form (1), written in SWRL on top of the ontology model. Our modeling approach included two types of rules: (a) intermediary rules to classify individuals to known profiles (i.e. composite classes) in terms of their attributes/responses and (b) main rules, which are the main knowledge content and able to indicate predictions or suggestions for specific cases, inferred from their classifications, properties and other parameters.

For example, a set of rules regarding hotel requirements is: (1) if the purpose of a trip is vacation and its duration is two weeks and the desirable cost of the hotel room is 50 to 100€, this trip is characterized as standard quality vacation and (2) if a trip is characterized as standard quality vacation and the visitor's age is 56-65, then the visitor requires security and cleanness. The first rule is an intermediate one and the second is a main rule. In SWRL they are written as:

   1:*Vacation(?mytrip*), hasTravelFeature *(?mytrip, 2WeekVisit*),
hasTravelFeature*(?mytrip, Cost50to100*) → *StandardQualityVacation(?mytrip)*

2:*StandardQualityVacationTrip(?mytrip)*, has Age*(?myvisitor, Age56-65*) →
hasRequirement*(?myvisitor, Security*), hasRequirement*(?myvisitor, Cleanness*)

Although the above syntax is difficult to read, the achievement is that the rules are machine understandable, they can be exchanged between systems and can be used to answer queries. The outcome of the rule formation process for the destination image survey was at the current stage 80 intermediate rules, 73 main rules and 19 analysis-based composite class definitions. It was encouraging that the introduction of findings from two independent surveys into a common Knowledge Base was performed successfully, taking advantage of the ability of the consolidated knowledge model to describe common concepts without inconsistencies, as well as to satisfy the requirements of both sources for terminology definitions and expressiveness.

4. Inference and Decision Support Mechanism

The knowledge content can be exploited by a marketer for decision support via a query mechanism. This mechanism is supported by an inference engine (or reasoner) which applies logic to produce inferred knowledge from the declared knowledge and to compute query results. The Protégé-OWL 4.2 environment includes preinstalled reasoner engines and tabs that present the inferred class hierarchy, the inferred class memberships and properties. It also offers a DL-Query tab for querying the ontology with OWL-DL expressions and a SPARQL tab which supports this more powerful RDF-based semantic web query language [32].

The use of these mechanisms to access the knowledge content is illustrated using a simple problem as an example. Suppose that we wish to study the visitors who come from Italy for swimming/seaside vacations and our target was the age category 56-65. We added in the ontology an individual named *myvisitor*, as member of class *VisitorWithDecisionPriority* which is a subclass of Visitor that is used to categorize visitors according to their decision factors. Then we asserted the following properties to myvisitor:

hasAge *Age56-65*
isFrom *Italy*
hasDecisionFactor *FirstPriorityBeaches*

After activating the reasoner, the above facts are exploited together with all axioms, relations and rules to infer additional properties that are true for myvisitor. In Figure 5, the description (i.e. class memberships) and the properties of myvisitor which have been explicitly given appear in boldface font, while the ones appearing in normal font and colored background fill are those that have been inferred by the reasoner. As shown in Figure 5, it comes out that myvisitor was classified as VisitorForSeaside and has the properties:

hasDecisionFactor *FirstPriorityTransports*
hasDecisionFactor *FirstPriorityNightlife*
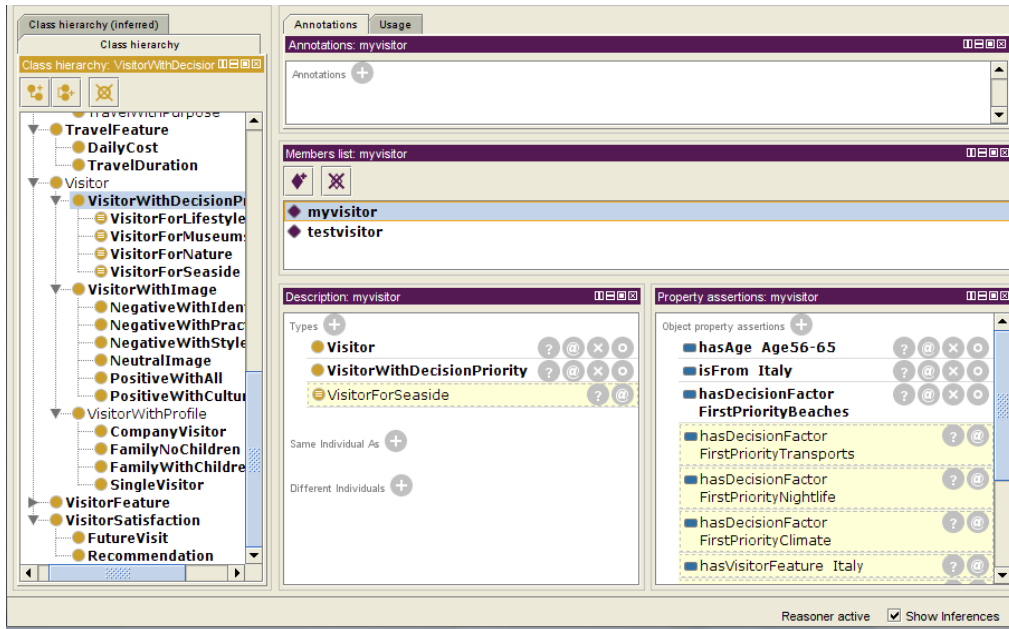hasDecisionFactor *FirstPriorityClimate*

Figure 5. Description and properties of the example individual visitor under study.

The system tells us e.g. that this type of visitor tends to consider seriously the quality of transports and is interested in the nightlife. By clicking on the "explain inference" button, we get the facts and rules that led to each conclusion. In Figure 6, we can see on what basis the system concluded that myvisitor has the property hasDecisionFactor *FirstPriorityTransports.*
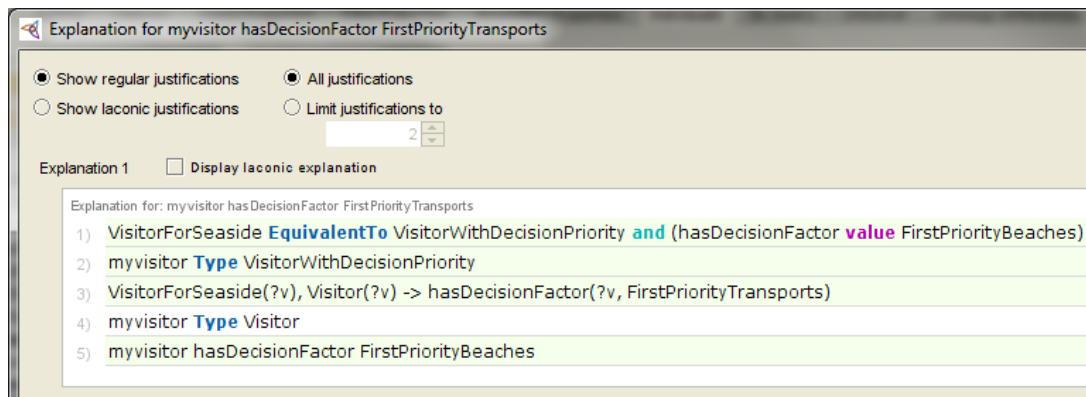

Figure 6. Explanation given by the system for the conclusion that the individual myvisitor having as decision factor the quality of transports.

The DL Query can be used to find which are the properties related to our target. The query: "DestinationDecisionFactor **and** (**inverse** (hasDecisionFactor) **value** myvisitor)" returns all the decision factors found to be true for myvisitor. In a similar way, additional queries can be applied to find which individuals have a specific property (e.g. what type of visitors are interested in low-budget vacation).
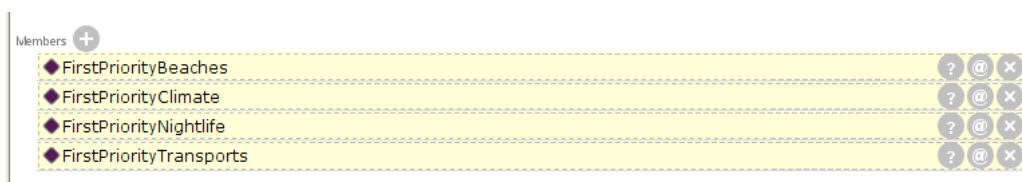

Figure 7. The result of the query that returns all the properties of the individual myvisitor.

5. Conclusions

In this paper, multidimensional data analysis methods were coupled with knowledge engineering technologies to develop a Knowledge-Based framework for extracting and managing knowledge for decision support in the destination marketing domain. A methodology was proposed for revealing interesting patterns for marketing and for expressing them in a structured, computerized form, so that they can be introduced in a knowledge-based decision support system. It was shown that the proposed analysis methods perform well as a knowledge extraction mechanism, having the ability to reveal qualitative information from survey data in explorative fashion.

In our pilot application, primary survey data on the image of Thessaloniki and the decision factors of its visitors were analyzed with advanced factor and clustering methods to uncover the dimensions in the way visitors see and select tourist destinations. Additionally, data on visitors' expectations and satisfaction from their hotel were analyzed to discover trends in their needs. In both surveys, classes were formed that reflected representative patterns of preferences and were associated to groups of visitors, thus resulting to marketing rules. The developed Knowledge Model was designed as a container of consolidated knowledge that resulted from different surveys, so that this knowledge can be maintained, exchanged and used for marketing decision support. An inference engine, the OWL Description Logic capabilities and the Protégé platform were used to exploit the knowledge content of the Knowledge Base.

Future steps include the addition of richer knowledge content and the performance of tests on the decision support abilities of the results in more complex tasks in realistic marketing projects. The populated KBS will then be tested regarding its problem solving abilities by the DMO of Thessaloniki in pilot marketing actions. An additional challenge is to introduce the time dimension into the model, in order to deal with temporal trends. Based on the feedback to be received, it is planned to eventually build a marketing platform offering solutions to typical marketing application scenario.

References

[1] Gunter, B. and Furnham, A. (1992), *Consumer Profiles: An Introduction to Psychographics*, Routledge, London.
[2] Weinstein, A. (1994). *Market Segmentation: Using Demographics, Psychographics, and other Niche Marketing Techniques to Predict and Model Customer Behaviour*, Probus, Chicago, IL.
[3] Etchner, C. and Ritchie, B. (1993). The measurement of destination image: an empirical assessment. *Journal of Travel Research*, pp. 3-13, Spring 1993.
[4] Cho, B. (1998). Assessing tourist satisfaction. *Tourism Recreation Research*, Vol. 23 No. 1, pp. 47-54.

[5] Kavoura, A. (2013). Politics of Heritage promotion: Branding the identity of the Greek state. *Tourism Culture and Communication*, 12, 69-83.

[6] Studer R., Benjamins R. and Fensel D. (1998). Knowledge Engineering: Principles and methods. *Data & Knowledge Engineering* 25 (1998) 161-197.

[7] Han J. and Kamber M., (2001). *Data Mining – Concepts and Techniques*, Academic Press, San Diego.

[8] Shadbolt N. and Milton, N. (1999). From knowledge engineering to knowledge management. *British Journal of Management*, vol. 10, 1999, 309–322.

[9] Schreiber, G. (2008). *Knowledge Engineering. In: Handbook of Knowledge Representation*, F. van Harmelen, V. Lifschitz, B. Porter (Eds.), Elsevier, 2008, pp. 929–946.

[10] Feigenbaum, E. A. and McCorduck, P. (1983). *The fifth generation* (1st ed.), Reading, MA: Addison-Wesley, ISBN 978-0-201-11519-2, OCLC 9324691

[11] Bousset, J.P., et al, (2007). A Decision Support System for Integrated Tourism Development: Rethinking Tourism Policies and Management Strategies. *Tourism Geographies*,9:4,387 — 404.

[12] Moutinho, L., Rita, P. and Curry, B. (1996). *Expert Systems in Tourism Marketing*, London and New York: Routledge.

[13] Benzecri, J-P, (1992). *Correspondence Analysis Handbook*. New-York: Dekker, P.

[14] Greenacre, M. (2007). *Correspondence Analysis in Practice*, Chapman & Hall.

[15] OWL, W3C Recommendation (2013). Available through the internet: http://www.w3.org/TR/owl-features/ [last visited: 30/5/2013].

[16] Information Systems Group (2014). HermiT OWL Reasoner. Available through the Internet: http://hermit-reasoner.com/ [last accessed 30/5/2014].

[17] Kollia, I., Glimm, B., Horrocks, I (2011). SPARQL Query Answering over OWL Ontologies. *Proc. 8th Extended Semantic Web Conf.* (ESWC'11).

[18] Stalidis G. and Karapistolis D. (2013). Multidimensional Data Analysis and Knowledge Management Technologies for Tourist Decision Support. *Proc of 3rd Intnl Conf Quantitative and Qualitative Methodologies in the Economic and Administrative Sciences (QMEAS)* Athens, 23-24 May, 2013, pp 373-379, ISBN: 978-960-98739-4-9.

[19] Stalidis, G. and Karapistolis, D. (2013). Tourist destination marketing supported by electronic capitalization of knowledge. *The 2nd International Conference on Strategic Innovative Marketing*, Prague 13-17 September, PROCEDIA Social and Behavioural Sciences (in press).

[20] Stalidis, G. (2013). Discovering marketing rules for the tourist sector in visitor service quality surveys. *Journal of Marketing Vistas*, Vol 3, No 2, pp 24-37.

[21] Van de Geer, John, P. (1993). *Multivariate Analysis of Categorical Data: Applications* Newbury Park: Sage Publications Inc. Advanced Quantitative Techniques in the Social Sciences Series Vol 3.

[22] Méthodes d' Analyses des Données - MAD (2012), Available through the internet: http://www.mkt.teithe.gr/en/research/research-labs/madlab/mad/, [last visited: 30/5/2014].

[23] Schreiber, G., Akkermans, H., Anjewierden, A., Hoog, R., Shadbolt, N., Van de Velde W. and Wielinga, B. (1999). *Knowledge Engineering and Management: The CommonKADS Methodology*. MIT Press, 1999.

[24] Protégé, (2014). Available through the internet: http://protege.stanford.edu/ [last visited:30/5/2014].

[25] Gruber,T.,(1993). A Translation Approach to Portable Ontology Specifications, *Knowledge Acquisition*, 5(2), pp. 199-220.

[26] Prantner, K., Ding, Y., Luger, M., Yan, Z. (2007). Tourism Ontology and Semantic Management System: State-of-the-Arts Analysis. *Proceedings of IADIS International Conference WWW/Internet 2007*, pp.111-115.

[27] Ou, S., Pekar, V., Orasan, C., Spurk, C., Negri M., (2008). Development and Alignment of a Domain-Specific Ontology for Question Answering, in European Language Resources Association (ELRA) (ed.): *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco.

[28]DERI, OnTour Ontology (2011). Available through the internet: http://etourism.deri.at/ont/index.html [last visited: 15/12/2011].

[29]SWRL: A Semantic Web Rule Language Combining OWL and RuleML (2004). W3C Member Submission 21 May 2004, National Research Council of Canada, Network Inference, and Stanford University. Available through the internet: http://www.w3.org/Submission/SWRL/ [last visited 30/5/2014].

[30]Segaran, Toby; Evans, Colin; Taylor, Jamie (2009). *Programming the Semantic Web*. O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472. p. 84. ISBN 978-0-596-15381-6.

[31]Ligêza, A., (2006). Logical foundations for rule-based systems. *Studies in Computational Intelligence*, vol. 11, 2nd Ed., Springer-Verlag Berlin Heidelberg.

[32]Prud'hommeaux, E., Seaborne, A. (eds.): SPARQL Query Language for RDF. W3C Recommendation (15 January 2008), Available through the internet: http://www.w3.org/TR/rdf-sparql-query/ [last visited 30/5/2014]