An AI Application to Integrated Tourism Planning

Francesca Alessandra Lisi(✉) and Floriana Esposito

Dipartimento di Informatica, Università degli Studi di Bari "Aldo Moro", Bari, Italy
{francesca.lisi,floriana.esposito}@uniba.it

Abstract. Integrated Tourism can be defined as the kind of tourism which is explicitly linked to the localities in which it takes place and, in practical terms, has clear connections with local resources, activities, products, production and service industries, and a participatory local community. In this paper we report our experience in applying Artificial Intelligence techniques to Integrated Tourism planning in urban areas. In particular, we have modeled a domain ontology for Integrated Tourism and developed an Information Extraction tool for populating the ontology with data automatically retrieved from the Web. Also, we have defined several Semantic Web Services on top of the ontology and applied a Machine Learning tool to better adapt the automated composition of these services to user demands. Use cases of the resulting service infrastructure are illustrated for the Apulia Region, Italy.

1 Introduction

Integrated Tourism can be defined as the kind of tourism which is explicitly linked to the localities in which it takes place and, in practical terms, has clear connections with local resources, activities, products, production and service industries, and a participatory local community. The goal of Integrated Tourism is twofold. For the various interests, requirements and needs the aim is to be fused together into a composite, integrated strategic tourism plan. For tourism the aim is to be planned with the intention of being fused into the social and economic life of a region and its communities. Although there is evidence that some tourism destinations have developed without conscious strategic and integrated planning, many of them have experienced unforeseen consequences (either physical, or human, or marketing or organizational impacts) which have led to their deterioration. Integrated Tourism has turned out to be crucial in the sustainable development of rural areas (so-called Integrated Rural Tourism) [19]. However, the integrated approach can be beneficial also to urban areas as testified by recent progress in *Urban Tourism* research [1]. Urban tourism is the practice of taking a vacation and visiting an inner-city area, as opposed to visiting such places as historical sites and natural features. Such cities market tourist attractions located within the city itself. This can include points of interest such as churches, a particular city block or feature, and amazing architectural creations. Urban Tourism is complex, difficult to pin down and define, and depends on

M. Gavanelli et al. (Eds.): AI*IA 2015, LNAI 9336, pp. 246–259, 2015.
DOI: 10.1007/978-3-319-24309-2_19

many factors such as the size of the town, its history and heritage, its morphology and its environment, its location, its image, etc. In the case of Integrated Urban Tourism, tourism is being seen as a cornerstone of a policy of urban development that combines a competitive supply able to meet visitors' expectations with a positive contribution to the development of towns and cities and the well-being of their residents.

In this paper, we report our experience in applying Artificial Intelligence (AI) techniques to Integrated Urban Tourism within the *Puglia@Service*[1] project. In particular, the techniques employed are borrowed from the AI areas of Knowledge Engineering (KE), Information Extraction (IE) and Machine Learning (ML). The resulting e-tourism application is intended to be compliant with the principles and the standards of the so-called *Semantic Web* [2]. The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries, mainly by means of shared vocabularies called *ontologies* [23]. In *Puglia@Service*, ontologies come into play in the definition of *Semantic Web Services* [9]. More precisely, Web Services are enriched with semantic annotations involving concepts taken from undelying domain ontologies. This will enable users and software agents to automatically discover, invoke, compose, and monitor Web resources offering services, under specified constraints, for Integrated Tourism in the Apulia Region, Italy.[2] Moreover, a distinguishing feature of the application described in this paper is that the automated composition of Semantic Web Services can be improved by learning from users' feedback. Preliminary and partial reports of the work done within *Puglia@Service* can be found in [12,14].

The paper is structured as follows. Section 2 provides some background information. In particular, Section 2.1 summarizes the goals of *Puglia@Service* whereas Section 2.2 reviews relevant literature on e-tourism. Section 3 shortly describes a domain ontology, named *OnTourism*, which models the relevant features of Integrated Tourism. Section 4 briefly presents a IE tool, named WIE-OnTour, which has been developed for populating *OnTourism* with data automatically retrieved from the Web. Section 5 illustrates some of the Semantic Web Services which have been defined on top of *OnTourism*. Section 6 introduces a ML tool, named Foil-$\mathcal{DL}$, which has been run on data provided by users in order to automatically generate new axioms for *OnTourism*. Section 7 concludes the paper with final remarks and directions of future work.

2 Background

2.1 Integrated Tourism Services in Apulia

Puglia@Service is an Italian PON Research & Competitivity project aimed at creating an innovative service infrastructure for the Apulia Region, Italy. The research conducted in *Puglia@Service* falls within the area of *Internet-based*

[1] http://www.ponrec.it/open-data/progetti/scheda-progetto?ProgettoID=5807
[2] http://en.wikipedia.org/wiki/Apulia

Service Engineering, *i.e.* it investigates methodologies for the design, development and deployment of innovative services. Concerning this area, the project will have an impact on the Apulia regional system at a strategic, organizational and technological level, with actions oriented to service innovation for the "sustainable knowledge society". The reference market of *Puglia@Service* is represented by the so-called *Knowledge Intensive Services* (KIS), an emerging category of the advanced tertiary sector, and transversal to the other economic sectors, that is supposed to play a prominent role within the restructuring process which will follow the world economic crisis.

Objective of the project is to promote a new service culture over the Apulia Region, marking a discontinuity point in the local development model, and guiding the transition of the region towards the "smart territory" paradigm where the territory is intended to be a multiplayer system able to improve, by means of an adequate technological and digital infrastructure, its attitude to innovation as well as its skills in managing the knowledge assets of the regional stakeholders. The arrangement of the new service model into the regional context will regard new generation services for Public Administration and for Integrated Tourism.

As for the application to Integrated Tourism (*Puglia@Service.Tourism*), which is of interest to this paper, the project addresses some of the issues analyzed in a report entitled "Sustainable Tourism and Local Development in Apulia Region" (2010)[3] and prepared by the Local Economic and Employment Development (LEED) Programme and the Tourism Committee of the Organisation for Economic Co-operation and Development (OECD) in collaboration with Apulia Region. The region as a touristic destination needs a better management in spite of the recent growth of visitors and the high potential. In particular, the report emphasizes the lack of an adequate ICT infrastructure and little use of new technologies.

Puglia@Service.Tourism encompasses an intervention on the Apulia tourism system, based on the definition of an Internet-based service model which increases the capability of KIS to create value for the region and for the tourist. Here, the tourist is not only "service user" but also "information supplier". In particular, the application will require the development of methods and technologies enabling an interaction model between the tourist and the territory with the ultimate goal of local development along three directions: Culture, Environment and Economy. For the purposes of this paper, we shall focus only on the Environment dimension along which *Puglia@Service.Tourism* aims at promoting forms of tourism with a low environmental impact centered around the notion of eco-compatible mobility. This will contribute to the achievement of a twofold goal. On one side, the tourist will benefit from decision support facilities during his/her tours, *e.g.* he/she will receive suggestions about sites of interest and public transportation means suitable to reach a certain destination. On the other side, the territory will benefit from the environmental sustainability of local tourism. The reduced environmental impact of eco-mobility together with the need for a more efficent transportation system in touristic places can be

[3] http://www.oecd.org/cfe/leed/46160531.pdf

obtained by combining sensoring tools and applications with rewarding mechanisms that encourage tourists and citizens to make eco-compatible choices. A possible scenario is described in the following. Once arrived in a touristic destination, the tourist could use his/her smartphone/PDA in order to obtain a suggestion about specific itineraries compliant with his/her profile and the information about the context. The tourist will be informed about the availabiliy of alternative transportation means and will be offered some credits for the green options (byking, trekking, car pooling, car sharing, etc.). In order to support this scenario, the *Puglia@Service.Tourism* infrastructure should deal with multi-dimensional information useful to suggest a touristic strategy which should meet users' expectations and preferences (in culture, enogastronomy, shopping, relax, etc.); environmental conditions, both meteorological and natural; multi-modal transportation means; availability of car pooling and car sharing services; transfer time between sites of interest. The "fingerprint" of tourists visiting an area in a given time span can be anonymized and employed to improve continuously the user profiling with the choices made by tourists with the same profile. To this aim it is necessary to track the trajectories of citizens and tourists by means of localization and wireless communication technologies (traces from mobile phones, PDA, vehicles with GPS, etc.).

It is straightforward to notice that Internet-based Service Engineering for KIS in Integrated Tourism should strongly rely on some Web technology enabling an automated service composition, just like Semantic Web Services.

2.2 State of the Art of e-Tourism

Most research on the application of Information and Communication Technology (ICT) to the tourism industry (so-called e-tourism) has been conducted by specializing technologies originally conceived for e-commerce (see [5] for a comprehensive yet not very recent review). However, e-tourism has been considered particularly challenging since the very beginning due to the technical issues raised by *interoperability*. Werthner and Klein [24] defined interoperability as the provision of a well-defined and end-to-end service which is in a consistent and predictable way. This generally covers not merely technical features but also in the case of electronic market environments, contractual features and a set of institutional rules. Interoperability enables partners to interact electronically with each other by the most convenient method and to deliver the right information at the right time to the right user at the right cost.

Ontologies have played a crucial role in solving the interoperability problem. Using a domain ontology a mediator software system (such as HARMONISE [7,18]) effectively "'translates"' partners' data and allows them to communicate electronically. Maedche and Staab [16,17] showed that Semantic Web technologies can be used for tourism applications to provide useful information on text and graphics, as well as generating a semantic description that is interpretable by machines. Dogac *et al.* [8] describe how to deploy semantically enriched travel Web services and how to exploit semantics through Web service registries.

They also address the need to use the semantics in discovering both Web services and Web service registries through peer-to-peer technology. Hepp *et al.* [10] investigate the use of ontological annotations in tourism applications. They show, based on a quantitative analysis of Web content about Austrian accommodations, that even a perfect annotation of existing Web content would not allow the vision of the Semantic Web to become a short-term reality for tourism-related e-commerce. Also, they discuss the implications of these findings for various types of e-commerce applications that rely on the extraction of information from existing Web resource, and stress the importance of Semantic Web Services technology for the Semantic Web. Siorpaes and Bachlechner [22] develop a system based on a fast and flexible Semantic Web backbone with a focus on e-tourism. The major benefits of this approach are its simplicity, modularity, and extensibility. In [11], Jakkilinki *et al.* describe the underlying structure and operation of a Semantic Web based intelligent tour planning tool. The proposed tour planner has inbuilt intelligence which allows it to generate travel plans by matching the traveller requirements and vendor offerings stored in conjunction with the travel ontology. Bousset *et al.* [3] present a decision support system which combines tools to assist in the analysis of the views, concerns and planned strategies of a wide range of tourism stakeholders in the face of given trends in tourists' expectations. Ricca *et al.* [21] present a successful application of logic programming for e-tourism: the iTravel system. The system exploits two technologies: (i) a system for ontology representation and reasoning, called OntoDLV; and, (ii) a semantic IE tool. The core of iTravel is an ontology which models the domain of tourism offers. The ontology is automatically populated by extracting the information contained in the tourism leaflets produced by tour operators. A set of specifically devised logic programs is used to reason on the information contained in the ontology for selecting the holiday packages that best fit the customer needs. An intuitive web-based user interface eases the task of interacting with the system for both the customers and the operators of a travel agency. Brilhante *et al.* [4] propose TripBuilder, a user-friendly and interactive system for planning a time-budgeted sightseeing tour of a city on the basis of the points of interest and the patterns of movements of tourists mined from user-contributed data. The knowledge needed to build the recommendation model is entirely extracted in an unsupervised way from two popular collaborative platforms: Wikipedia and Flickr. TripBuilder interacts with the user by means of a friendly Web interface that allows her to easily specify personal interests and time budget. The sightseeing tour proposed can be then explored and modified.

3 Modeling Domain Knowledge in *OnTourism*

Modeling the knowledge concerning the domain of Integrated Tourism is the first of the KE activities we have performed within *Puglia@Service.Tourism*. Domain ontologies for tourism are already available, *e.g.* the *travel*[4] ontology is centered around the concept of *Destination*. However, it is not fully satisfactory from the

[4] http://www.protege.cim3.net/file/pub/ontologies/travel/travel.owl

viewpoint of Integrated Tourism. For instance, it lacks concepts modeling the reachability of places. Therefore, we have decided to build a domain ontology, named *OnTourism*,[5] more suitable for the project objectives and compliant with the OWL2 standard.[6] It consists of 379 axioms, 205 logical axioms, 117 classes, 9 object properties, and 14 data properties, and has the DL expressivity of $\mathcal{ALCOF}(\mathbf{D})$ according to Protégé [7] ontology metrics. [8]

The main concepts forming the terminology of *OnTourism* model the sites (class *Site*), the places (class *Place*), and the distances between sites (class *Distance*). Sites of interest include accommodations (class *Accommodation* with subclasses *Bed_and_Breakfast* and *Hotel* among the others), attractions (class *Attraction* with several subclasses, *e.g.*, *Church*), stations (class *Station*), and civic facilities (class *Civic*). The terminology encompasses the amenities (class *Amenity* with subclasses like *Wheelchair_Access*) and the services (class *Service* further specialized in, *e.g.*, *Car_Rental* and *Bike_Rental*) offered by hotels. Also, it models the official 5-star classification system for hotel ranking (class *Rank* with instances *1_star*, *2_stars*, and so on) as well as a user classification system for accommodation rating (class *Rate* with instances *Excellent*, *Very_Good*, *Average*, *Poor*, *Terrible*). Finally, the terminology includes landscape varieties (class *Landscape* with instances *City*, *Country*, *Lake*, *Mountain*, *River*, and *Sea*) and transportation means (class *Transportation_Mean* with instances *Bike*, *Car*, *Foot*, and *Public_transit*). Distances are further classified into *Distance_by_car* and *Distance_on_foot* according to the transportation means used.

The object properties in *OnTourism* model the relationship between a site and a distance (*hasDistance*), the relationship between a distance and the two sites (*isDistanceFor*), and the relationship between a site and the place where the site is located at (*isLocatedAt*) (see Figure 1). Also, for each accommodation, it is possible to specify the amenities available (*hasAmenity*) and the services provided (*provides*). The user rating allows to classify accommodations into five categories (from *Excellent_Accommodation* to *Terrible_Accommodation*). In the case of hotels, the ranking (*hasRank*) is the starting point for the definition of five categories (from *Hotel_1_Star* to *Hotel_5_Stars*).

The data properties in *OnTourism* allow to refer to sites by *name* and to places by *address*, *zipcode*, *city*, and *country*. Details about accommodations are the number of rooms (*numberOfRooms*) and the average price of a room (*hasPrice*). Distances between sites have a numerical value in either length or time units (*hasLengthValue*/*hasTimeValue*). Note that each of these numerical values would be better modeled as attribute of a ternary relation. However, only binary relations can be represented in OWL. The concept *Distance* and the properties *hasDistance*, *isDistanceFor* and *hasLengthValue*/*hasTimeValue* are necessary to simulate a ternary relation by means of binary relations.

[5] http://www.di.uniba.it/~lisi/ontologies/OnTourism.owl
[6] http://www.w3.org/TR/owl2-overview/
[7] http://protege.stanford.edu/
[8] *OnTourism* significantly extends the *Hotel* ontology (described in [15]).

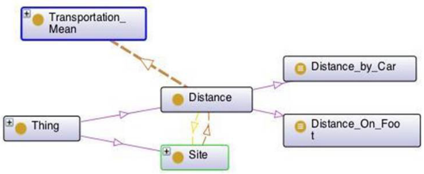


Fig. 1. Portion of the *OnTourism* ontology modeling the distances between sites.

4 Populating *OnTourism* with WIE-OnTour

At the ontology population step, the ontology is augmented with instances of concepts and properties. This KE activity is often perfomed by relying on some IE tool. IE aims at the automated extraction of structured information from unstructured and/or semi-structured machine-readable documents. In most of the cases this activity concerns processing texts by means of natural language processing. The proliferation of the Web has intensified the need for developing IE systems that help people to cope with the enormous amount of data that is available online, thus giving raise to Web Information Extraction (WIE) [6]. WIE tools typically exploit the HTML/XML tags and layout format that are available in online text. As a result, less linguistically intensive approaches have been developed for IE on the Web using wrappers, which are sets of highly accurate rules that extract a particular page's content. Wrappers typically handle highly structured collections of web pages, such as product catalogues and telephone directories. They fail, however, when the text type is less structured, which is also common on the Web.

WIE-OnTour is a wrapper-based WIE tool implemented in Java and conceived for the population of *OnTourism* with data concerning accommodations (in particular, those in the categories "Hotel" and "B&B") available in the web site of TripAdvisor[9]. Note that the tool does not generate instances of the classes *Amenity* and *Service* because it is not necessary for our purposes. However, amenities and services can be implicitely declared. For instance, the fact that a hotel, say *hotel_10*, offers a bike rental service is modeled by means of an axiom stating that *hotel_10* is instance of the class $\exists hasAmenity.Bike\_Rental$. The tool is also able to compute distances of the extracted accommodations from sites of interest (*e.g.*, touristic attractions) by means of Google Maps[10] API. Finally, the tool supports the user in the specification of sites of interest. A snapshot of WIE-OnTour is shown in Figure 2.

Instantiations of *OnTourism* for the main destinations of urban tourism in Apulia have been obtained with WIE-OnTour. Here, we consider an instantiation for the city of Bari (the capital town of Apulia). It contains 34 hotels, 70 B&Bs, 106 places, 208 distances for a total of 440 individuals. Information about

[9] http://www.tripadvisor.com/
[10] http://maps.google.com/

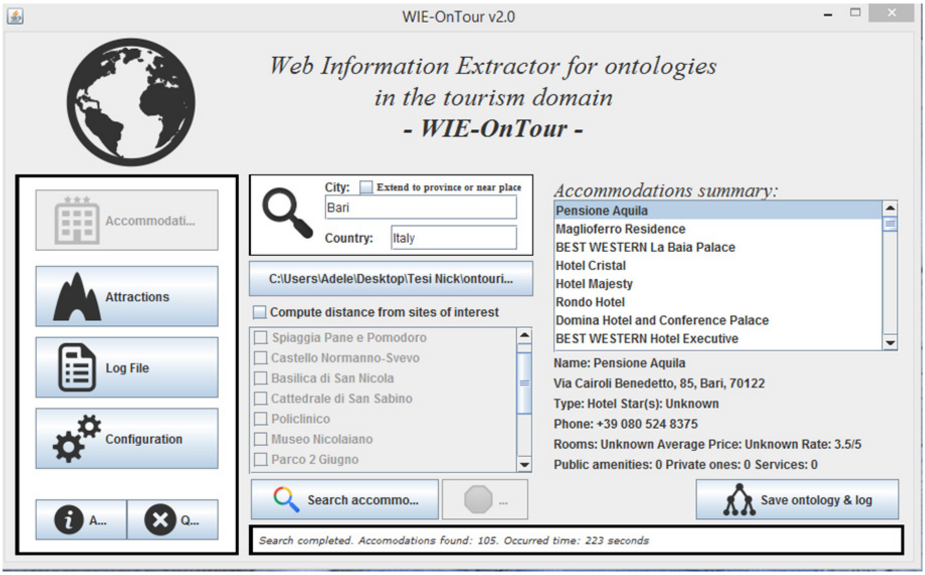


Fig. 2. Web Information Extraction for the city of Bari, Italy, with WIE-OnTour.

the rank, the amenities, the services, and the average room price has been added in the ontology for each extracted accommodation. The distances are provided in time and length on foot and have been computed with respect to *Basilica di San Nicola*[11] and *Cattedrale di San Sabino*[12] (both instances of *Church*). The restriction to foot distances is due to the aforementioned preference of Integrated Tourism for eco-mobility.

5 Building Semantic Web Services Upon *OnTourism*

Semantic Web Services are the server end of a client-server system for machine-to-machine interaction via the Web [9]. As a component of the Semantic Web, they are defined with mark-up languages which make data machine-readable in a detailed and sophisticated way. In particular, *OWL-S*[13] is an ontology which provides a standard vocabulary that can be used together with the other aspects of OWL to create service descriptions. The use of OWL-S makes it easy for programmers to combine data from different sources and services without losing meaning. Web services can be activated "behind the scenes" when a Web browser makes a request to a Web server, which then uses various Web services to construct a more sophisticated reply than it would have been able to do on its own. Semantic Web Services can also be used by automatic programs that run

[11] Basilica of St. Nicholas: http://en.wikipedia.org/wiki/Basilica_di_San_Nicola
[12] Cathedral of St. Sabinus: http://en.wikipedia.org/wiki/Bari_Cathedral
[13] http://www.w3.org/Submission/OWL-S/

without any connection to a Web browser. Overall, the interchange of semantic data allows to overcome some of the limits of conventional Web services. Indeed, the mainstream XML standards for interoperation of Web services specify only syntactic interoperability, not the semantic meaning of messages. For example, *Web Services Description Language* (WSDL)[14] can specify the operations available through a Web service and the structure of data sent and received but can not specify the semantic meaning of the data or semantic constraints on the data. This requires programmers to reach specific agreements on the interaction of Web services and makes automatic Web service composition difficult.

In *Puglia@Service.Tourism*, we have defined several services on top of *OnTourism* (and *travel*) by following the OWL-S approach. For example, let us consider *destination_attractions_service* that returns the attractions located in a given destination. In OWL-S it can be semantically described as an atomic service with only one input and only one output where the parameter types for the input and the output are the classes *Destination* (belonging to *travel*) and *Attraction* (occurring in *OnTourism*) respectively. Several specializations of *destination_attractions_service* have been considered, one for each subclass of the parameter types. For example, *city_churches_service* returns the churches (output parameter of type *Church*) located in a given city (input parameter of type *City*). When executed for the city of, *e.g.*, Bari, the service will query the underlying domain ontologies (more precisely, their instance level) to retrieve each *Church* that *isLocatedAt* some *Place* in Bari, *e.g. Basilica di San Nicola* and *Cattedrale di San Sabino*. Note that these instances will be returned also by *destination_attractions_service* because they are inferred to be instances of *Attraction*. As a further case, *near_attraction_accomodations_service* returns all the accommodations (output parameter of type *Accommodation*) near a given attraction (input parameter of type *Attraction*). Note that closeness can be defined on the basis of the distance between sites (class *Distance*) either in a crisp way (*i.e.*, when the distance value is under a fixed threshold) or in a fuzzy way (*i.e.*, through grades of closeness). In both ways, however, the computation should consider the transportation means used (*Distance_by_car* vs. *Distance_on_foot*) as well as the measure units adopted (*hasLengthValue* vs. *hasTimeValue*).

In *Puglia@Service.Tourism*, we have chosen to define only OWL-S atomic services in order to exploit the notorious advantages of the WSDL grounding. Composite services can be automatically obtained by applying service composition methods such as the one described in [20]. The simplest form of composite service is based on the control construct of *Sequence*. For example, the services *city_churches_service* and *near_attraction_accomodations_service* can be executed in sequence by having the output of the former as input to the latter. Note that the type mismatch is only apparent since *Church* is a subclass of *Attraction*. One such service composition satisfies, *e.g.*, the user request of knowing the accommodations around *Basilica di San Nicola* and *Cattedrale di San Sabino* in Bari. Considering that Bari is a major destination of religious tourism in Apulia, this composite service effectively supports the demand from pilgrims who prefer to

[14] http://www.w3.org/TR/2007/REC-wsdl20-primer-20070626/

find an accommodation in the neighborhood of places of worship so that they can practise their own religions at any hour of the day. Also, if the suggested accommodations are easy to reach (*i.e.*, at foot distance) from the site of interest, the service will bring benefit also to the city, by reducing the car traffic. In a more complex scenario, disabled pilgrims might need a wheelchair-accessible accommodation. The service composition mechanism should then append a further specialized service, say *wheelchairaccess_hotels_service*, which returns the hotels (output parameter of type *Hotel*) with disabled facilities (input parameter of type *Wheelchair_Access*). Indeed, the resulting composite service can be considered more compatible with the special needs of this user profile.

6 Improving Services by Learning from Users' Feedback

In *Puglia@Service.Tourism*, the automated composition of Semantic Web Services will be enhanced by exploiting users' feedback. The idea is to apply ML tools in order to induce ontology axioms which can be used for discarding those compositions that do not reflect the preferences/expectations/needs of a certain user profile. Here, we illustrate this idea with an application scenario which builds upon the accommodation rating provided by TripAdvisor's users. More precisely, we consider the task of accommodation finding which consists of distinguishing good accommodations from bad ones according to the amenities available, the services offered, the location, and the distance from sites of interest. In order to address this classification problem, we need ML tools able to deal with the inherent incompleteness of Web data and the inherent vagueness of concepts such as closeness. One such tool is FOIL-$\mathcal{DL}$.

FOIL-$\mathcal{DL}$ implements a method for learning a set of *fuzzy* General Concept Inclusion (GCI) axioms from positive and negative examples for a target class A_t in any *crisp* OWL 2 ontology [15]. In particular, the GCIs inducible with FOIL-$\mathcal{DL}$ are $\mathcal{EL}(\mathbf{D})$ axioms of the form

$$C \sqsubseteq A_t \ , \tag{1}$$

where the left-hand side is defined according to the following syntax

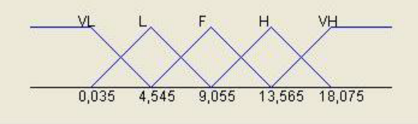
$$C \longrightarrow \top \mid A \mid \exists R.C \mid \exists T.\mathbf{d} \mid C_1 \sqcap C_2 \ . \tag{2}$$

Here, A is a primitive concept, R is an object property, T is a data property, C, C_1, C_2 are complex concepts, and $\mathbf{d}$ can be any of the following membership functions widely used in fuzzy set theory

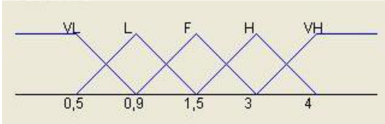
$$\mathbf{d} := ls(a,b) \mid rs(a,b) \mid tri(a,b,c) \ , \tag{3}$$

where *e.g.* $ls(a,b)$ is the left-shoulder membership function.

For instance, the 5 fuzzy sets that can be automatically generated by FOIL-$\mathcal{DL}$ for the data property *hasLengthValue* of *OnTourism* are graphically shown in Figure 3. Here, we have used the membership functions $ls(a,b)$ for the first



(a)



(b)

Fig. 3. Fuzzy sets (*VeryLow* (VL), *Low* (L), *Fair* (F), *High* (H), and *VeryHigh* (VH)) derived from the concrete domain used as range of the data property *hasLengthValue* of *OnTourism* in the case of (a) *Distance_by_car* and (b) *Distance_on_foot*.

fuzzy set, $tri(a, b, c)$ for the subsequent three ones, and $rs(a, b)$ for the fifth one. Also, we have distinguised between the two cases of *Distance_by_car* and *Distance_on_foot*. Notably, in the latter case, a very low distance does not exceed 900 meters, an average distance is about 1500 meters, and so on. New concepts such as *hasLengthValue_VeryLow* are obtained from these fuzzy sets and can appear in the GCI axioms induced by FOIL-$\mathcal{DL}$. For example, the axiom

$$Hotel\_3\_Stars \sqcap \exists hasDistance.(\exists isDistanceFor.(Train\_Station) \sqcap hasLengthValue\_VeryLow) \sqcap hasPrice\_Fair \sqsubseteq Good\_Hotel$$

translates into the following sentence in natural language: ‘A 3-star hotel with a very low distance from a train station and a fair room price is a good hotel’.

More formally, given a crisp $\mathcal{DL}$[15] ontology $\mathcal{K}$ (*background theory*), an atomic concept A_t (*target concept*), a set $\mathcal{E} = \mathcal{E}^+ \cup \mathcal{E}^-$ of *crisp* concept assertions labelled as either positive or negative examples for A_t (*training set*), and a set $\mathcal{L_H}$ of *fuzzy* $\mathcal{EL}(\mathbf{D})$ GCI axioms (the *language of hypotheses*) of the form (1), the goal of FOIL-$\mathcal{DL}$ is to find a set $\mathcal{H} \subset \mathcal{L_H}$ (a *hypothesis*) of axioms satisfying the following properties: (i) $\forall e \in \mathcal{E}^+, \mathcal{K} \cup \mathcal{H} \models e$ (*completeness*), and (ii) $\forall e \in \mathcal{E}^-, \mathcal{K} \cup \mathcal{H} \not\models e$ (*consistency*). Further details about FOIL-$\mathcal{DL}$ are reported in [15].

As an illustration of the potential usefulness of FOIL-$\mathcal{DL}$ in Integrated Tourism planning, we discuss here a couple of experiments concerning the filtering of results returned by the services reported in Section 5. We set up a learning problem with the class *Bad_Accommodation* (abbr. *BA*) as target concept. Ratings from TripAdvisor users have been exploited for distinguishing good accommodations from bad ones by using WIE-ONTOUR (see Section 4). Out of the 104 accommodations, 57 with a higher percentage (say, over 70%) of positive users’ feedback are asserted to be instances of *Good_Accommodation*, whereas 15 with a lower percentage (say, under 50%) are asserted to be instances of *Bad_Accommodation*. The latter, of course, play the role of positive examples in our learning problem. Syntactic restrictions are imposed on the form of the learnable GCI axioms. More precisely, conjunctions can have at most 3 conjuncts and at most 2 levels of nesting are allowed in existential role restrictions. The two trials differ as for the alphabet underlying the language of hypotheses.

[15] $\mathcal{DL}$ stands for any DL.

In the first experiment, we have not considered the distances of the accommodations from the sites of interest (*i.e.*, we have forbidden the use of *hasDistance* in $\mathcal{L}_{\mathcal{H}}$). With this configuration, FOIL-$\mathcal{DL}$ returns just the following GCI with confidence 0.5:

$$B\&B \sqcap \exists hasAmenity.(Pets\_Allowed) \sqcap \exists hasAmenity.(Wheelchair\_Access) \sqsubseteq BA$$

The GCI suggests that B&Bs are not recommended even though they provide disabled facilities. It can be used to filter out from the result set of *wheelchairaccess_accommodations_service* those accommodations which are classified as bad.

In the second experiment, conversely, we have considered the distances of the accommodations from the sites of interest (*i.e.*, we have enabled the use of *hasDistance* in $\mathcal{L}_{\mathcal{H}}$). With this configuration, FOIL-$\mathcal{DL}$ returns the following GCI with confidence 1.0:

$$\exists hasAmenity.(Bar) \sqcap \exists hasAmenity.(Wheelchair\_Access) \sqcap \\ \exists hasDistance.(\exists isDistanceFor.(B\&B) \sqcap \exists isDistanceFor.(Church)) \sqsubseteq BA$$

The GCI strenghtens the opinion that B&Bs are not recommendable accommodations for disabled people whatever their distance from the churches is.

As a further experiment, we have restricted our analysis of accommodations in Bari to only B&Bs. Starting from 12 positive examples and 39 negative examples for *Bad_Accommodation*, FOIL-$\mathcal{DL}$ returns the following two GCIs with confidence 0.154 and 0.067 respectively:

$$\exists hasAmenity.(Pets\_Allowed) \sqcap \exists hasAmenity.(Wheelchair\_Access) \sqsubseteq BA \\ \exists hasAmenity.(Bar) \sqcap \exists hasAmenity.(Wheelchair\_Access) \sqsubseteq BA$$

which confirm that B&Bs should not be recommended to disabled tourists.

7 Conclusions and Future Work

In this paper we have reported our ongoing work on the use of AI techniques for supporting Integrated Tourism services in the urban areas of the Apulia region within the *Puglia@Service* project. More precisely, we have shortly described *OnTourism*, a domain ontology for Integrated Tourism. Also, we have briefly presented WIE-ONTOUR, a WIE tool which has been developed for populating *OnTourism* with data automatically retrieved from the Web sites of TripAdvisor and Google Maps. Moreover, we have illustrated the semantic descriptions in OWL-S of some Integrated Tourism services built on top of *OnTourism*. Finally, we have outlined an application scenario for FOIL-$\mathcal{DL}$, a ML tool able to deal with incomplete data and vague concepts, whose findings can be used to enhance the automated composition of OWL-S services.

Though developed for the purposes of the project, the technical solutions here described are nevertheless general enough to be reusable for similar applications in other geographical contexts. Overall, the application shows the added value of having ontologies and ontology reasoning behind an Interned-based service infrastructure. Note that by ontology reasoning we intend also non-standard forms such as the inductive inference supported by FOIL-$\mathcal{DL}$. The use of ML algorithms in our application makes a great difference from the related works mentioned in Section 2.2, including iTravel [21] which is the closest in design and implementation. However, at the current stage, the application suffers from some limits. First, WIE-ONTOUR is hard-coded. More precisely, the tool needs to be continuously maintained to keep up with updates of the structure of Web sites which are inspected for extracting the data. Second, FOIL-$\mathcal{DL}$ can not deal with big data. Tests over larger cities show that the system does not scale well.

For the future we intend to carry on the work on the application of FOIL-$\mathcal{DL}$ to the automated service composition. Notably, we shall consider the problem of learning from the feedback provided by specific user profiles. Also, in order to overcome the limits of FOIL-$\mathcal{DL}$, we are investigating the possibility of reducing the size of data by means of fuzzy granulation [13]. This should have a positive impact especially on the number of instances for the distance relation, and consequently on the computational cost of the learning process.

Acknowledgments. This work was partially funded by the Italian PON R&C 2007-2013 project PON02_00563_3489339 "*Puglia@Service*: Internet-based Service Engineering enabling Smart Territory structural development".

References

1. Ashworth, G., Page, S.J.: Urban tourism research: Recent progress and current paradoxes. Tourism Management **32**, 1–15 (2011)
2. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American, May 2001
3. Bousset, J.P., Skuras, D., Tešitel, J., Marsat, J.B., Petrou, A., Fiallo-Pantziou, E., Kušová, D., Bartoš, M.: A decision support system for integrated tourism development: Rethinking tourism policies and management strategies. Tourism Geographies **9**(4), 387–404 (2007)
4. Brilhante, I., Macedo, J.A., Nardini, F.M., Perego, R., Renso, C.: Tripbuilder: a tool for recommending sightseeing tours. In: de Rijke, M., Kenter, T., de Vries, A.P., Zhai, C.X., de Jong, F., Radinsky, K., Hofmann, K. (eds.) ECIR 2014. LNCS, vol. 8416, pp. 771–774. Springer, Heidelberg (2014)
5. Buhalis, D., Law, R.: Progress in information technology and tourism management: 20 years on and 10 years after the internet - the state of eTourism research. Tourism Management **29**(4), 609–623 (2008)
6. Chang, C.H., Kayed, M., Girgis, M.R., Shaalan, K.F.: A survey of web information extraction systems. IEEE Transactions on Knowledge and Data Engineering **18**(10), 1411–1428 (2006)
7. Dell'Erba, M., Fodor, O., Höpken, W., Werthner, H.: Exploiting semantic web technologies for harmonizing e-markets. J. of IT & Tourism **7**(3–4), 201–219 (2005)
8. Dogac, A., Kabak, Y., Laleci, G., Sinir, S.S., Yildiz, A., Kirbas, S., Gurcan, Y.: Semantically enriched web services for the travel industry. SIGMOD Record **33**(3), 21–27 (2004)
9. Fensel, F., Facca, F.M., Simperl, E., Toma, L.: Semantic Web Service. Springer, Heidelberg (2011)

10. Hepp, M., Siorpaes, K., Bachlechner, D.: Towards the semantic web in e-tourism: can annotation do the trick? In: Ljungberg, J., Andersson, M. (eds.) Proceedings of the Fourteenth European Conference on Information Systems, ECIS 2006, Göteborg, Sweden, pp. 2362–2373 (2006)
11. Jakkilinki, R., Georgievski, M., Sharda, N.: Connecting destinations with an ontology-based e-tourism planner. In: Sigala, M., Mich, L., Murphy, J. (eds.) Information and Communication Technologies in Tourism, ENTER 2007, Proceedings of the International Conference in Ljubljana, Slovenia, pp. 21–32. Springer (2007)
12. Lisi, F.A., Esposito, F.: Semantic web services for integrated tourism in the apulia region. In: Giordano, L., Gliozzi, V., Pozzato, G.L. (eds.) Proceedings of the 29th Italian Conference on Computational Logic, Torino, Italy, June 16–18, 2014. CEUR Workshop Proceedings, vol. 1195, pp. 178–193. CEUR-WS.org (2014)
13. Lisi, F.A., Mencar, C.: Towards fuzzy granulation in OWL ontologies. In: Ancona, D., Maratea, M., Mascardi, V. (eds.) Proceedings of the 30th Italian Conference on Computational Logic, Genova, Italy, June 30 – July 2, 2015. CEUR Workshop Proceedings. CEUR-WS.org (2015)
14. Lisi, F.A., Esposito, F.: Supporting integrated tourism services with semantic technologies and machine learning. In: Horridge, M., Rospocher, M., van Ossenbruggen, J. (eds.) Proceedings of the ISWC 2014 Posters & Demonstrations Track a Track within the 13th International Semantic Web Conference, ISWC 2014, Riva del Garda, Italy, October 21, 2014. CEUR Workshop Proceedings, vol. 1272, pp. 361–364. CEUR-WS.org (2014)
15. Lisi, F.A., Straccia, U.: A FOIL-like method for learning under incompleteness and vagueness. In: Zaverucha, G., Santos Costa, V., Paes, A. (eds.) ILP 2013. LNCS, vol. 8812, pp. 123–139. Springer, Heidelberg (2014)
16. Mädche, A., Staab, S.: Applying semantic web technologies for tourism information systems. In: Wöber, K., Frew, A., Hitz, M. (eds.) Information and Communication Technologies in Tourism 2002, pp. 311–319. Springer, Vienna (2002)
17. Mädche, A., Staab, S.: Services on the move - towards P2P-enabled semantic web services. In: Frew, A., Hitz, M., O'Connor, P. (eds.) Information and Communication Technologies in Tourism 2003. Springer (2003)
18. Missikoff, M., Werthner, H., Hoepken, W., Dell'Erba, M., Fodor, O., Formica, A., Taglino, F.: Harmonise - towards interoperability in the tourism domain. In: Frew, A., Hitz, M., O'Connor, P. (eds.) Information and Communication Technologies in Tourism 2003, pp. 58–66. Springer, Vienna (2003)
19. Oliver, T., Jenkins, T.: Sustaining rural landscapes: The role of integrated tourism. Landscape Research **28**(3), 293–307 (2003)
20. Redavid, D., Iannone, L., Payne, T.R., Semeraro, G.: OWL-S atomic services composition with SWRL rules. In: An, A., Matwin, S., Raś, Z.W., Ślezak, D. (eds.) Foundations of Intelligent Systems. LNCS (LNAI), vol. 4994, pp. 605–611. Springer, Heidelberg (2008)
21. Ricca, F., Dimasi, A., Grasso, G., Ielpa, S.M., Iiritano, S., Manna, M., Leone, N.: A logic-based system for e-tourism. Fundamenta Informaticae **105**(1–2), 35–55 (2010)
22. Siorpaes, K., Bachlechner, D.: Ontour: tourism information retrieval based on YARS. In: Demos and Posters of the 3rd European Semantic Web Conference (ESWC 2006), Budva, Montenegro, 11th 14th June, 2006 (2006)
23. Staab, S., Studer, R. (eds.): Handbook on Ontologies, 2nd edn., International Handbooks on Information Systems. Springer (2006)
24. Werthner, H., Klein, S.: Information Technology and Tourism - A Challenging Relationship. Springer Verlag, Wien (1999)