Supporting Integrated Tourism Services with Semantic Technologies and Machine Learning

Francesca A. Lisi and Floriana Esposito

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

Abstract. In this paper we report our ongoing work on the application of semantic technologies and machine learning to Integrated Tourism in the Apulia Region, Italy, within the *Puglia@Service* project.

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. Integrated Tourism thus needs ICTs that should go beyond the mere technological support for tourism marketing, differently from most approaches in eTourism research (see [1] for a comprehensive yet not very recent review). In this paper, we report our experience in supporting Integrated Tourism services with Semantic Technologies (STs) and Machine Learning (ML). The work has been conducted within *Puglia@Service*,[1] an Italian PON Research & Competitivity project aimed at creating an innovative service infrastructure for the Apulia Region, Italy.

The paper is structured as follows. Section 2.1 shortly describes a domain ontology for Integrated Tourism, named *OnTourism*, which has been modeled for being used in *Puglia@Service*. Section 2.2 briefly presents a Web Information Extraction (WIE) tool, named WIE-OnTour, which has been developed for populating *OnTourism* with data automatically retrieved from the Web. Section 2.3 illustrates some of the Semantic Web Services (SWSes) which have been defined on top of *OnTourism* for supporting Integrated Tourism in Apulia. Section 3 outlines an application scenario for a ML tool, named Foil-$\mathcal{DL}$, to better adapt the automated composition of these services to user demands. Section 4 concludes the paper with final remarks and directions of future work.

2 Semantic Technologies for Integrated Tourism

2.1 A Domain Ontology

Domain ontologies for tourism are already available, *e.g.* the *travel* [2] ontology is centered around the concept of *Destination*. However, it is not fully satisfactory

[1] http://www.ponrec.it/open-data/progetti/scheda-progetto?ProgettoID=5807
[2] http://www.protege.cim3.net/file/pub/ontologies/travel/travel.owl

from the viewpoint of Integrated Tourism because, *e.g.*, it lacks concepts modeling the reachability of places. In *Puglia@Service*, we have decided to build a domain ontology, named *OnTourism*, [3] more suitable for the project objectives and compliant with the OWL 2 standard. It consists of 379 axioms, 205 logical axioms, 117 classes, 9 object properties, and 14 data properties, and has the expressivity of the DL $\mathcal{ALCOF}(\mathbf{D})$.

The main classes of the terminology are *Site*, *Place* and *Distance*. The first is the root of a taxonomy which covers several types of sites of interest (*e.g.*, *Hotel* and *Church*). The second models the places where sites are located at. The third, together with the object properties *hasDistance* and *isDistanceFor* and the data properties *hasLengthValue*/*hasTimeValue*, allows to represent the distance relation between sites with values in either length or time units. Distances are further specified according to the transportation means used (see, *e.g.*, the class *Distance_on_Foot*). Other relevant classes in the terminology are *Amenity* (with subclasses such as *Wheelchair_Access*) and *Service* (with subclasses such as *Bike_Rental*) that model, respectively, amenities and services available at the accommodations. Finally, the terminology includes the official 5-star classification system for hotel ranking.

2.2 Ontology Population with Web Information Extraction

WIE-OnTour is a wrapper-based WIE tool implemented in Java and conceived for the population of *OnTourism* with data concerning hotels and B&Bs available in the web site of TripAdvisor[4]. The tool is also able to compute distances of the extracted accommodations from sites of interest (*e.g.*, touristic attractions) by means of the Google Maps[5] API. Finally, the tool supports the user in the specification of sites of interest.

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, and 208 foot distances for a total of 440 individuals. The distances are provided in time and length on foot and have been computed with respect to *Basilica di San Nicola* and *Cattedrale di San Sabino* (both instances of *Church* and located in Bari). The restriction to foot distances is due to the aforementioned preference of Integrated Tourism for eco-mobility.

2.3 Semantic Web Services

In *Puglia@Service*, we have defined several atomic services in OWL-S on top of the aforementioned domain ontologies, *travel* and *OnTourism*. For example, *city_churches_service* returns the churches (o.p. of type *Church*) located in a given city (i.p. of type *City*) whereas *near_attraction_accomodations_service* returns all the accommodations (o.p. of type *Accommodation*) near a given attraction (i.p.

[3] `http://www.di.uniba.it/~lisi/ontologies/OnTourism.owl`
[4] `http://www.tripadvisor.com/`
[5] `http://maps.google.com/`

of type *Attraction*). Note that closeness can be defined on the basis of 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 as well as the measure units adopted according to the *OnTourism* ontology.

In *Puglia@Service*, we intend to obtain composite services by applying methods such as [3]. For example, the sequence composed of *city_churches_service* and *near_attraction_accomodations_service* could satisfy, *e.g.*, the user request for accommodations around *Basilica di San Nicola*. Indeed, since Bari is a major destination of religious tourism in Apulia, it could effectively support the demand from pilgrims who prefer to 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 also *wheelchairaccess_accommodations_service*, so that the resulting composite service could be considered more compatible with the special needs of this user profile.

3 Towards Learning from Users' Feedback

In *Puglia@Service*, automated service composition 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. This task strongly relies on a classification problem aimed at distinguishing good accommodations from bad ones according to the amenities available, the services offered, the location and the distance from sites of interest, etc. 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 the closeness. One such tool is Foil-$\mathcal{DL}$ [2], a ML system able to induce a set of fuzzy General Concept Inclusion (GCI) $\mathcal{EL}(\mathbf{D})$ axioms from positive and negative examples for a target class in any OWL ontology.

As an illustration of the potential usefulness of Foil-$\mathcal{DL}$ in the *Puglia@Service* context, we report here a couple of experiments concerning the filtering of results returned by the SWSes reported in the previous section for the case of Bari. We set up a learning problem with the class *Bad_Accommodation* as target of the learning process. Ratings from TripAdvisor users have been exploited for providing Foil-$\mathcal{DL}$ with positive and negative examples. Out of the 104 accommodations, 57 with a higher percentage (say, over 0.7) of positive users' feedback are asserted as instances of *Good_Accommodation*, whereas 15 with a lower percentage (say, under 0.5) are asserted as 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.

In the first experiment, we have not considered the distances of the accommodations from the sites of interest. With this configuration, FOIL-$\mathcal{DL}$ returns just the following GCI with confidence 0.5:

```
Bed_and_Breakfast and hasAmenity some (Pets_Allowed) and hasAmenity some (Wheelchair_Access)
        subclass of Bad_Accommodation
```

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. With this configuration, FOIL-$\mathcal{DL}$ returns the following GCI with confidence 1.0

```
hasAmenity some (Bar) and hasAmenity some (Wheelchair_Access) and
hasDistance some (isDistanceFor some (Bed_and_Breakfast) and isDistanceFor some (Church))
    subclass of Bad_Accommodation
```

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:

```
hasAmenity some (Pets_Allowed) and hasAmenity some (Wheelchair_Access) subclass of Bad_Accommodation
hasAmenity some (Bar) and hasAmenity some (Wheelchair_Access) subclass of Bad_Accommodation
```

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

4 Conclusions and future work

In this paper we have reported our ongoing work on the use of STs and ML for Integrated Tourism in Apulia within the *Puglia@Service* project. 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. Notably, they show the added value of having ontologies and ontology reasoning (including also non-standard inferences like induction as exemplified by FOIL-$\mathcal{DL}$) behind a Web Service infrastructure.

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.

References

1. 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)
2. Lisi, F.A., Straccia, U.: A System for Learning GCI Axioms in Fuzzy Description Logics. In: Eiter, T., Glimm, B., Kazakov, Y., Kroetzsch, M. (eds.) Proc. of the 26th Int. Workshop on Description Logics. CEUR Workshop Proceedings, vol. 1014. CEUR-WS.org (2013)
3. Redavid, D., Iannone, L., Payne, T.R., Semeraro, G.: OWL-S atomic services composition with SWRL rules. In: An, A., Matwin, S., Ras, Z.W., Slezak, D. (eds.) Foundations of Intelligent Systems. LNCS, vol. 4994, pp. 605–611. Springer (2008)