Engineering a Comprehensive Romanian Tourism Ontology

Adrian GROZA*, Radu Razvan SLAVESCU*, Irina DRAGOSTE*, Bernadette VARGA†
*Intelligent Systems Group,
Department of Computer Science,
Technical University of Cluj-Napoca, Romania
Adrian.Groza@cs.utcluj.ro
†Recognos Romania SRL
Bernadette.Varga@recognos.ro

Abstract — **This paper presents a large domain ontology for the Romanian tourism. The ontology was developed in the KRSS syntax in a modular way. After introducing the core ontology, we detail various models that encapsulate knowledge on: accommodation types and their facilities, touristic activities, points of interest or eating and drinking resorts. For populating the ontology various sources were used: Foursquare, Open Street Map or the Open Linked Data for Romania. The ontology was validated against a set of competency questions formalized in nRQL.**

I. INTRODUCTION

The ontology was developed in RacerPro (Renamed ABox and Concept Expression Reasoner Professional) knowledge representation and reasoning system [2]. In our view, RacerPro and the corresponding Knowledge Representation System Specification (KRSS) syntax for Description Logic axioms are powerful technical instrumentation that support ontology engineering behind the basic capabilities provided by GUI-based ontology editors. Specifically, the KRSS syntax facilitates the speed of writing axioms, while RacerPro provides a wide set of primitives to introduce concepts, roles, constraints, debugging axioms and query the knowledge base.

II. DESCRIPTION LOGIC IN KRSS SYNTAX

In the description logic *ALC*, concepts are built using the set of constructors formed by negation, conjunction, disjunction, value restriction, and existential restriction [1], as shown in Table I. Here, *C* and *D* represent concept descriptions, while *r* is a role name. The semantics is defined based on an interpretation $I = (\Delta^I, \cdot^I)$, where the domain Δ^I of I contains a non-empty set of individuals, and the interpretation function $\cdot^I$ maps each concept name

C to a set of individuals $C^I \in \Delta^I$ and each role *r* to a binary relation $r^I \in \Delta^I \times \Delta^I$. The last column of Table I shows

the extension of for non-atomic concepts.

An ontology consists of terminologies (or TBoxes) and assertions (or ABoxes). A terminology *TBox* is a finite set of terminological axioms of the form (*equiv C D*) or (*implies C D*). An assertional box *ABox* is a finite set of concept assertions (*instance a C*) or role assertions (*related a b r*), where *C* designates a concept, *r* a role, and *a* and *b* are two individuals. Usually, the unique name assumption holds within the same *ABox*. A concept *C* is satisfied if

there exists an interpretation *I* such that $C^I \neq \emptyset$. The concept

D subsumes the concept *C*, represented by (*implies C D*) if

$C^I \subseteq D^I$ for all interpretations *I*. Constraints on concepts

(i.e. *disjoint*) or on roles (*domain*, *range* of a role, *inverse* roles, or *transitive* properties) can be specified in more expressive description logics[2]

TABLE I. KRSS SYNTAX AND SEMANTICS OF *ALC*.

Constructor	*Syntax*	*Semantics*
negation	`(not C)`	$\Delta^I \setminus C^I$
conjunction	`(and C D)`	$C^I \cap D^I$
disjunction	`(or C D)`	$C^I \cup D^I$

This work was supported by the Romania-Moldova Bilateral Agreement entitled "ASDEC: Structural Argumentation for Decision Support with Normative Constraints", and by the PN-II Innovation Checks Programme UEFSCDI Romania for supporting innovation in small and medium enterprises, through the project "LELA - Collaborative Recommendation System in the Tourism Domain Using Semantic Web Technologies and Text Analysis in Romanian Language". A. Groza is supported by the UTCN through the internal research project GREEN-VANETS.

[2] We provide only some basic terminologies of description logics in this paper to make it self-contained. For a detailed explanation about families of description logics, the reader is referred to [1].

existential restriction	`(some r C)`	$\{x \in \Delta^I \mid \exists y : (x,y) \in r^I \wedge y \in C^I\}$
value restriction	`(all r C)`	$\{x \in \Delta^I \mid \forall y : (x,y) \in r^I \rightarrow y \in C^I\}$
individual assertion	`(instance a C)`	$\{a\} \in C^I$
role assertion	`(related a b r)`	$(a^I, b^I) \in r^I$

III. Reusing related ontologies

The process of engineering the ontology has started by: a) specifying use cases of the ontology, b) defining a set of competency questions, and c) analyzing the existing ontologies for possible reuse.

a) Use cases: The main *use case* of the ontology is to create a guide application for young tourists visiting Cluj-Napoca. A usage scenario would be: *"Many young tourists will visit Cluj-Napoca, the Youth European Capital in 2015. With an expected average stay of 4 days, they want to see as much and variate as possible"*.

b) Competency questions: A solution to narrow the scope of an ontology is to start by defining a list of *competency questions* (CQs) [3]. CQs are questions that an ontology should be able to answer in order to satisfy use cases. Thereby, CQs represent initial requirements and they can be used to validate the ontology. Having the role of a requirement, each CQs are written in natural language (see Table II). For the validation task, the CQs are formalized in nRQL (new Racer Query Language) [2].

c) Reusing ontologies: As many tourism ontologies do exist, we employ the domain coverage metrics to select the

TABLE II. Sample of competency questions for a tourism ontology.

CQ1	What services are included in a specific accommodation?
CQ2	What time is check-in/out for a given accommodation?
CQ3	What places to eat and drink are within a given distance?
CQ4	What points of interest are around the accommodation?
CQ5	What activities can you do around the accommodation?
CQ6	Which are the traveling options around a point of interest (POI)

most adequate ontologies. The domain coverage metric is based on the semantic similarity between concepts the concepts represented by ontology classes and the ones described by user-given search terms. Our solution [6] counts the concepts that completely match a term or one of its synonyms. This value is normalized by the number of terms given. The synonyms of the terms, in parentheses, are selected from WordNet suggested synonyms, according to the intended word meaning. The results of domain coverage calculation, using our proposed class matching method, is presented in

Assume that the following terms and their synonyms were specified: *cruise (sail), mountain (mount), monument (memorial), museum, traveling, camping (tenting, encampment), hiking (tramp)*. Given a repository of tourism ontologies (like swoogle.com), the domain coverage computed for each ontology against the above search terms is depicted in Table III. The ontologies with the highest score were analyzed for possible reuse of various concepts and roles.

IV. Engineering the ontology

To develop the tourism ontology, we follow the methodology in [3] and we also enact various ontology design patterns [4]. The ontology is a modular one, consisting of a core formalization and T-boxes for modeling various aspects in the tourism domain.

A. Core ontology

Fig.1 illustrates the main four classes of the Romanian tourism ontology (*Accommodation*, *POI*, *Gastro*, *Activity*) and the main relationship between them: *hasActivity*, *hasAccommodation*, *hasEatingAndDrinking* and *hasPointOfInterest.* The spatial location is attached through the *hasLocation* role between the four concepts and the *Location* concept.

TABLE III. Domain coverage for 17 ontologies in the tourism domain.

Ontology Id	Ontology URI	Domain Coverage
102	http://rewerse.net/A1/otn/OTN.owl	0.2857
103	http://harmonisa.uni-klu.ac.at/ontology/skeleton.owl	0.0
104	http://www.info.uqam.ca/Members/valtchev_p/mbox/ETP-tourism.owl	0.1429
105	http://harmonisa.uni-klu.ac.at/ontology/moland.owl	0.1429
106	http://fivo.cyf-kr.edu.pl/ontologies/test/VOTours/TravelOntology.owl	0.1429
107	http://cui.unige.ch/isi/onto/2010/urba-en.owl	0.5714
108	http://en.openei.org/wiki/Special:ExportRDF/South_Africa_Department of Environment Affairs and Tourism	0.0
109	http://en.openei.org/wiki/Special:ExportRDF/Climate_Change_Adaptation and Mitigation in the Tourism Sector	0.0
111	http://jxml2owl.projects.semwebcentral.org/sample/tourism.owl	0.0
112	http://iri.columbia.edu/~benno/data_center.owl	0.0
113	http://www.pms.ifi.lmu.de/rewerse-wga1/otn/OTN.owl	0.2857
114	http://aabs-semanticweb-prototypes.googlecode.com/svn-history/r2/trunk/ontologies/2007/02/Test/needs.rdf	0.0
115	http://aabs-semanticweb-prototypes.googlecode.com/svn-history/r2/trunk/ontologies/2007/02/Flight/Flight.owl	0.0
116	http://aabs-semanticweb-prototypes.googlecode.com/svn-history/r2/trunk/ontologies/2007/02/Places/Places.owl	0.1429
117	http://www.esd.org.uk/standards/lgcl/1.03/lgcl-schema/lgcl.xml	0.0
118	http://www.cs.ox.ac.uk/isg/ontologies/lib/GardinerCorpus/http___protege.stanford.edu_plugins_owl_owl-library_travel.owl/2009-02-13/00120.owl	0.1429
119	http://harmonisa.uni-klu.ac.at/ontology/realraum.owl	0.0

```
(equivalent TourismAxis (or Accommodation Activity
                            Gastro POI Location))
(disjoint Accommodation Activity
          Gastro POI Location)
(define-role hasAccommodation
    :domain (or Activity Gastro POI)
    :range   Accommodation)
(define-role hasActivity
    :domain (or Accommodation Gastro POI)
    :range   Activity)
(define-role hasEatDrink
    :domain (or Accommodation Activity POI)
    :range  Gastro)
(define-role hasPOI
    :domain (or Accommodation Activity Gastro)
    :range POI)
(define-role hasLocation :domain TourismAxis
                         :range Location
                         :transitive t
                         :inverse LocatedIn)
```

FIG. 1. TOP LEVEL CONCERNS IN THE TOURISM ONTOLOGY

Part of the ontology is automatically populated with information extracted from tourism blogs. The aim is to extract subjective impressions about named entities in our domain. To model this, we introduced the concept Match for representing the relations between the touristic places and the blog posts where the named entity was identified. This concept was modelled by enacting the n-ary ontology design pattern [5]. The goal was to combine several information about a tourism blog regarding: subject of the blog according to the concepts in the ontology, computed score about an instance in the ontology, or provenance information like author, starting and ending text index which relates to an individual in the ontology.

As an example, the individual *m1* of type *Match* is related to the blog *b100* via the role *fromBlogPost*. The point of interest *mateicorvin* is related to the same match *m1* by the relation *hasSubject*. The positive score of 0.8 in line 22 is computed with a basic opinion mining algorithm from the blog post.

```
(define-role fromBlogPost :domain Match
                          :range BlogPost)
(define-role hasSubject :domain Match
                        :range TourismAxis)
(define-concrete-domain-attribute hasScore
    :domain Match :type real)
(define-concrete-domain-attribute hasText
    :domain Match :type string)
(define-role speaksAbout :domain BlogPost
                         :range TourismAxis)

(instance m1 Match)
(instance b100 BlogPost)
(instance mateicorvin POI)
(attribute-filler m1
   "casa matei corvin atrage multi turisti" hasText)
(related m1 b100 fromBlogPost)
(related m1 mateicorvin hasSubject)
(attribute-filler m1 0.8)
```

FIG. 2. RELATING INFORMATION ABOUT A BLOG WITH THE N-ARY DESIGN PATTERN.

B. Extended T-boxes of the ontology

This subsection details the extension of the core ontology with knowledge related to accommodation and points of interest. The extended ontology is modular.

d) Modeling knowledge on accommodation: Each accommodation provides various facilities. The corresponding TBox contains a list of 200 facilities. Fig. 3 presents 20 of these facilities.

```
(implies (or InternetAccess AirConditioning Golf
             Restaurant Babysitting Gym-fitness
             Casino KidsClub Beachaccess Sofabed
             Washingmachine Parking Refrigerator
             Iron Changeofthelinen Terrace-Balcony
             Diningroom Weeklycleaning)
         Faccility)
```

FIG. 3. SAMPLE FROM THE 200 FACILITIES FORMALIZED IN THE *LELA* ONTOLOGY.

Each accommodation type is located in a city. The ontology uses an ABox of 428 cities in Romania. We use two ABoxes for asserting facts about accommodation. The first ABox includes 555 individuals of type *Accommodation*. These instances are categorized in the following categories: *GoodHotel*, *LuxuryHotel*, *ExtraLuxuryHotel* or *Budget* accommodation, according to the number of stars (see Fig. 4).

A booking is encapsulated in the ontology as:

```
(instance b1 Booking
          (string= has-start-date "2014-05-13")
          (string= has-end-date "2014-06-15")
          (= has-price 130.25))
```

The second ABox contains 2517 hotels, with related information described by the following features: *name*, *has-address*, *has-phone*, *has-website* (see Fig. 5).

e) Modelling touristic activities: We consider several types of tourism (i.e., *agriculture*, *nautical*, *medical*, *culinary*, *popCulture*, *cultural*, *extreme*, *heritage*, *warTourism*, *wellness*, *wildlife*) following the pattern:

```
(implies extremeTourism (and tourism
   (some based visitingDangerousPlaces)))
```

We also defined 61 touristic-related activities. Fig. 6 lists a sample of 20 such activities. The ontology contains a taxonomy of 80 geographical touristic objectives, following the pattern (*instance |Cascada Rachitele| |Waterfall|*).

Information about administrative areas is stored in terms of *region*, *country*, *city* and 462 *communes*.

f) Modelling points of interest: The tourism ontology includes relevant information about 637 museums in Romania.

Each museum has a textual description in Romanian, using the role has-description. Additionally, each museum has the following features: *has-location*, has-*email*, *has-phone*, *has-founding-year*, *has-latitude*, *has-longitude*, *county-has-museum*, *has-schedule*, *has-website*.

We focuses also on modelling points of interests (i.e., the ontology contains data on 953 caves in Romania).

```
(implies GoodHotel
  (and Hotel (has-value hasRating ThreeStarRating)))
(implies LuxuryHotel
  (and Hotel (has-value hasRating FourStarRating)))
(implies ExtraLuxuryHotel
  (and Hotel (has-value hasRating FiveStarRating)))
(implies Campground
  (and Accomodation
       (has-value hasRating OneStarRating)))
(implies Hostel (and Accommodation
                     (all hasRating TwoStarRating)))
(implies Chalet (and Accommodation
                     (all hasRating ThreeStarRating)
                         ))
(implies BudgetAccommodation (and Accommodation
   (some hasRating (or OneStarRating
                       TwoStarRating))))
```

FIG. 4. FORMALIZING HOTEL CLASSIFICATIONS.

```
(instance |Hotel2517| |Hotel|)
(attribute-filler |Hotel2517| "HotelHotel Zimbru"
    name)
(instance |CaleaBucovinei 1-3, CampulungMoldovenesc,
     RO| |Address|)
(related |Hotel2517| |CaleaBucovinei 1-3,
    CampulungMoldovenesc, RO| has-address)
(related |Hotel2517| |+40230314356| has-phone)
(related |Hotel2517| |www.suceava360.ro/|
    has-website)
```

FIG. 5. THE SECOND ABOX FOR ASSERTING INFORMATION ABOUT 2517 HOTELS IN ROMANIA.

```
(implies (or Basketball BeachVolleyball Biking
         BootCamp Calisthenics ChoppingFirewood
         CircuitTraining DanceRevolution
         EllipticalTrainer FamilyHike
         FieldHockey FlagFootball Football
         HalfMarathon Handball Hockey
         IceHockey IceSkating InlineSkating
         IrishDancing) Activity)
```

FIG. 6. TOURISTIC ACTIVITIES.

```
(instance |Muntii Bihorului| |Mountain|)
(related |ALBA| |Muntii Bihorului| has-mountain)
```

FIG. 7. ROMANIAN MOUNTAINS IN THE TOURISM ONTOLOGY.

Each cave has a specific location, as exemplified by the following RacerPro code:

```
(related |CLUJ| |Pestera Zanelor| has-cave)
```

The concept *Mountain* is instantiating with the corresponding individuals in Romania (see Fig. 7) In the same figure, each mountain belongs to a particular county.

The ontology contains also 729 instances of the concept *POI*.

C. Populating the ontology.

Data from Open Street Map (OSM) can be directly integrated as an ABox in the tourism ontology. For converting OSM into KRSS syntax we developed a java-based converter based on the OSMOSIS API.

The Foursquare taxonomy and individuals are also imported in the ontology. As an example, the concepts in the ArtsEntertainment domain from the Foursquare are listed in Fig. 8.

```
(implies Aquarium ArtsEntertainment)
(implies Arcade ArtsEntertainment)
(implies Art_Gallery ArtsEntertainment)
(implies Bowling_Alley ArtsEntertainment)
(implies Casino ArtsEntertainment)
(implies Comedy_Club ArtsEntertainment)
(implies Concert_Hall ArtsEntertainment)
```

FIG. 8. IMPORTING FOURSQUARE TAXONOMY IN OUT ONTOLOGY.

```
(define-primitive-role hasCuisineQuality :domain
    Gastro :range Match)
(instance m12345 Match)
(instance b101 BlogPost)
(related m12345 b101 fromBlogPost)
(attribute-filler m12345 "at NapoliCentrale one can
    eat well" hasText)
(attribute-filler m12345 1.0 hasScore)
(related pizzeriaNapoliCentrale m12345
    hasCuisineQuality)
```

FIG. 9. POPULATING THE ONTOLOGY THROUGH NATURAL LANGUAGE PROCESSING AND SENTIMENT ANALYSIS OF TOURISM BLOGS.

V. RETRIEVING INFORMATION FROM ABOXES

Assume that the blog 101 states that “at NapoliCentrale one can eat well”: the information obtained after performing sentiment analysis on the blog is asserted in the ontology as described by Fig 9.

The following listing illustrates three operations: i) checking the ontology consistency, ii) retrieve information about individuals in the ontology and iii) identify the sub-concepts of the main five axis of the ontology.

```
(tbox-cyclic?)
(tbox-coherent?)
(describe-individual m12345)
(concept-instances Location)
(concept-instances Activity)
(concept-instances POI)
(concept-instances Gastro)
(concept-instances Accommodation)
```

Various competency questions were formalized in nRQL. To obtain all activities and their locations, one can use:

```
(retrieve (?activity ?location)
     (and (?activity Activity)
          (?activity ?location hasLocation)))
```

To obtain all eating and drinking options and their locations, the following query can be enacted:

```
(retrieve (?ed) (and (?ed Gastro)
                     (?ed ?location hasLocation)))
```

To list all touristic objective with a positive review score greater than 0.8, we can use the RacerPro command:

```
(concept-instances (and Match (>= hasScore 0.8)))
```

To enumerate all points of interest and activities for which the location is explicitly specified we can use:

```
(retrieve (?x) (?x (has-known-successor
    POI_has_Location)))
(retrieve (?x) (?x  (has-known-successor
    Activity_has_Location)))
```

VI. CONCLUSION

To our knowledge, this the most comprehensive ontology for the Romanian tourism. The ontology was developed to be used in an industrial application [7] for the Recognos company.

REFERENCES

[1] F. Baader, *The description logic handbook: theory, implementation, and applications*. Cambridge university press, 2003.

[2] V. Haarslev, K. Hidde, R. Moller, and M. Wessel, “The RacerPro knowledge representation and reasoning system,” *Semantic Web*, vol. 3,no. 3, pp. 267–277, 2012.

[3] N. F. Noy, D. L. McGuinness *et a*l., “Ontology development 101: A guide to creating your first ontology,” 2001.

[4] J. T. Pollock and R. Hodgson, “Ontology design patterns,” *Adaptive Information: Improving Business through Semantic Interoperability, Grid Computing, and Enterprise Integration*, pp. 145–194.

[5] V. Presutti and A. Gangemi, “Content ontology design patterns as practical building blocks for web ontologies,” in *Conceptual Modeling-ER*. Springer, 2008, pp. 128–141.

[6] A. Groza, I. Dragoste, I. Sincai and I. Jimborean. An ontology selection and ranking system based on analytical hierarchy process, 16th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, Timisoara, Romania, 22-25 September 2014.

[7] B. Varga, A. Trambitas-Miron, A. Roth, A. Marginean, R. Slavescu, A. Groza - A natural language processing system for Romanian tourism, ASIR@ Fedcsis., Warsaw, Polonia, 7-10 September, 2014.