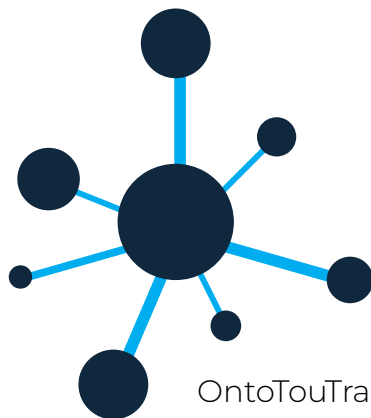# OntoTouTra: Tourist Traceability Ontology Based on Big Data Analytics

Ontology Implementation - Supplementary Material

Juan Francisco Mendoza-Moreno, Luz Santamaria-Granados,
Anabel Fraga Vázquez, Gustavo Ramirez-Gonzalez

November 12, 2021

OntoTouTra

Abstract

Tourist traceability is the analysis of the set of actions, procedures, and technical measures that allows us to identify and record the time-space causality of the tourist's touring, from the beginning to the end of the chain of the tourist product. Besides, the traceability of tourists has implications for infrastructure, transport, products, marketing, the commercial viability of the industry, the management of the destination's social, environmental, and cultural impact. To this end, a tourist traceability system requires a knowledge base for processing elements, such as functions, objects, events, and logical connectors between them. A knowledge base provides us with information on preparation, planning, and implementation or operation stages. In this regard, unifying tourism terminology in a traceability system is a challenge because we need a central repository that promotes standards for tourists and suppliers in forming a formal body of knowledge representation. Some studies are related to the construction of ontologies in tourism, but none focuses on tourist traceability systems. For the above, we are proposed OntoTouTra, an ontology that uses formal specifications to represent knowledge of tourist traceability systems. This document is the compendium of SPARQL queries to implement the OntoTouTra conceptual test cases proposed in the design paper of this ontology.

Contents

List of Figures

List of Tables

Introduction

This document relates the SPARQL queries and their results that we apply to the OntoTouTra ontology to validate it from the conceptual point of view from two approaches: The data-oriented one and the test of the ontology through competency questions. We also describe the tools we use to perform these queries, either directly through SPARQL from an endpoint or through the REST API.

This document is supplementary material to the paper: "OntoTouTra: Tourist Traceability Ontology based on Big Data Analytics." A set of ten test cases was configured, one for each KPI or CQ chosen for the ontology domain.

Chapter 1

OntoTouTra conceptual evaluation

To validate the efficiency of OntoTouTra, we created a set of tests to verify the conceptual model and a use case. This validation is based on two approaches: The data-driven [1], where real situations from the ontology domain are represented, and in the second approach, the ontology test [2], we answer Competency Questions (CQ) formulated by domain experts.

In the case of the OntoTouTra data-oriented test, we created individuals gathered through Web Scraping from an OTA (Booking.com). We collected data on the destination, accommodations, services, experiences, ratings, and reviews. We chose Colombia as a use case. Therefore we filtered the data from the OTA through this country.

Regarding the set of competence questions formulated by experts in the field of tourism, we used as sources the UNWTO and ONS, and data providers of the tourism sector in Colombia, such as the Ministry of Industry, Commerce and Tourism and platforms as SITUR. The bank of KPI of the ONS [3] has four categories or boxes: Satisfaction, Economy, sustainability, and Organizational. We chose and adapted ten KPI from this bank closely related to OntoTouTra's domain: tourism traceability. For each KPI, we elaborated a test case, implementing the respective SPARQL query. After an extensive review of sources from the tourism experts and authorities noted above, the key competency questions of the ontology were specified as follows:

- KPI01: What percentage of visitors are satisfied with the provider's services?

- KPI02: What percentage of users are satisfied with the provider's internet services?

- KPI03: Number of daily visitors.

- KPI04: Impact on the destination of the offer of accommodation companies used by visitors.

- KPI05: Impact of visits on the destination.

- KPI06: Influence of accommodation companies in the destination.

- KPI07: Arrival of foreign tourists (FTA)

- KPI08: Inbound and local tourism.

- KPI09: Seasonality patterns in the destination.

- KPI10: Portfolio of tourist experiences used.

Subsequently, to solve the question of each KPI, we elaborated the SPARQL query and executed it on an OntoTouTra endpoint, and the results of these consulted were compared with the data obtained from the source of the domain expert. In this way, we demonstrated the effectiveness of ontology from a conceptual point of view.

Table 1.1 depicts the test cases for each of the selected KPI. As a reference for comparison, local government and UNWTO sources were sought to contrast the expected results. The test cases were run using SPARQL queries whose results demonstrated the reliability of the ontology when compared with the expected results.

Table 1.1: Expected results

Test case	KPI	Expected results	Comparison sources	Source's data	Results obtained	Note
T001	1	Over 60 % of visitors rated the experience as good or excellent (see Listing 2.1)		-	71.56%	
T002	2	In Colombia, over 50% of customers consider the WiFi service to be good or excellent (see Listing 2.2)		-	53.5%	
T003	3	In Colombia, in 2019, over 1000 reviews per day (see Listing 2.3)	Colombia's Fact Sheets [1], pages 1-2	4,100,000 annual (2019)	2,423 (mean)	Booking's reviewers represent the 21.57% visitors
T004	4	In Colombia, two (2) accommodation enterprises per 10000 population (see Listing 2.4)	Colombia's Fact Sheets [1], page 4	5,6	2.33	28,000 establishments / 50 millon inhabitants = 5,6. Booking = 2,33
T005	5	The number of reviews depends on the local tourism industry (33 departments in Colombia) (see Listing 2.5)	Colombia's Tourism Report [2], page 18	Bogotá, Antioquia, Bolívar	Bogotá, Antioquia, Bolívar	Top-3 departments
T006	6	Population rate with hotel influence depends of the local tourism industry (see Listing 2.6)	Colombia's Tourism Report [2], page 28	San Andrés, Bolívar, Bogotá	Bogotá, San Andrés, Valle	Top-3 departments
T007	7	Top 10 Foreign Tourist Arrivals (FTAs) in Colombia (see Listing 2.7)	Colombia's Tourism Report [2], page 7	USA, Peru, France	USA, France, Argentina	Top-3 countries
T008	8	Inbound and domestic tourism in Colombia per Department (see Listing 2.8)	Colombia's Fact Sheets [1], pages 1-2	4,100,000	459,322	Inbound travels
T009	9	Seasonality Patterns per month of 2019 in Colombia (see Listing 2.9)	UNWTO Seasonality [3]	Jan-Mar, Jul-Aug	Jan-Apr, Jul-Aug	Peak seasons
T010	10	Top 10 Tourist experiences in Colombia (see Listing 2.10)		-	Beach, Tours, Game room	Top-3 tourist experiences

---

[1] UNWTO - Country Fact Sheets: https://webunwto.s3.eu-west-1.amazonaws.com/s3fs-public/2020-10/colombia.pdf

[2] MinComercio Colombia - Tourism Report - December, 2019: https://www.mincit.gov.co/estudios-economicos/estadisticas-e-informes/informes-de-turismo

[3] Dashboard on Tourism Seasonality: https://www.unwto.org/seasonality

Chapter 2

Test cases

2.1 Test case 1: What percentage of visitors are satisfied with the provider's services?

Listing 2.1: KPI-01 % of visitors who rate the overall visitor experience as good or excellent

```
PREFIX xsd:  <http://www.w3.org/2001/XMLSchema#>
PREFIX ott:  <http://tourdata.org/ontotoutra/ontotoutra.owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ((COUNT(?goodratings)) / (COUNT(?allratings)) AS ?percentage)
WHERE {
  {
     ?review ott:hotelReviewRating ?allratings
  }
  UNION
  {
    ?review ott:hotelReviewRating ?goodratings
    FILTER(?goodratings >= 8) }
  }
```

Result: **"0.71557521825831204326234"^^xsd:decimal**

Interpretation: 71.5% of the reviews were rated greater than or equal to 8.0 (Good).

2.2 Test case 2: What percentage of users are satisfied with the provider's internet services?

Listing 2.2: KPI-02 % of customers who consider the overall impression of the WiFi service to be good or excellent

```
PREFIX xsd:  <http://www.w3.org/2001/XMLSchema#>
PREFIX ott:  <http://tourdata.org/ontotoutra/ontotoutra.owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT ((COUNT(?score)) / (COUNT(?wifiscores)) AS ?percentage)
WHERE {
{
  ?category     rdf:type ott:ScoreCategory                   ;
                ott:scoreCategoryDescription "Free WiFi"@en .
  ?wifiscores   ott:hasScoreCategory ?category               .
}
UNION
{
  ?category     rdf:type ott:ScoreCategory                   ;
                ott:scoreCategoryDescription "Free WiFi"@en .
  ?hotelScore   ott:hasScoreCategory ?category               ;
                ott:score ?score                             .
  FILTER(?score >= 8)
}
}
```

Result: **"0.535060294774452880750335"^^xsd:decimal**

Interpretation: 53.5% of customers consider the WiFi service as good or excellent (≥ 8).

2.3 Test case 3: Number of daily visitors

Listing 2.3: KPI-03 Number of day visitors (Visitors who reviewed) frame

```
PREFIX xsd:  <http://www.w3.org/2001/XMLSchema#>
PREFIX ott:  <http://tourdata.org/ontotoutra/ontotoutra.owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT ?date (COUNT(?user) AS ?visitors)
WHERE {
  ?review rdf:type ott:HotelReview  ;
                   ott:hotelReviewDate ?date ;
                   ott:hotelReviewUser ?user .
  FILTER(?date > "2019-01-01T00:00:00"^^xsd:dateTime)
}
GROUP BY ?date
ORDER BY ASC(?date)
LIMIT 10
```

Expert: UNWTO

COLOMBIA

NOTE: Please interpret with caution. For the full data set, including metadata and footnotes, please refer to the
UNWTO Database and the Methodological Notes to the UNWTO Database, available through the UNWTO website
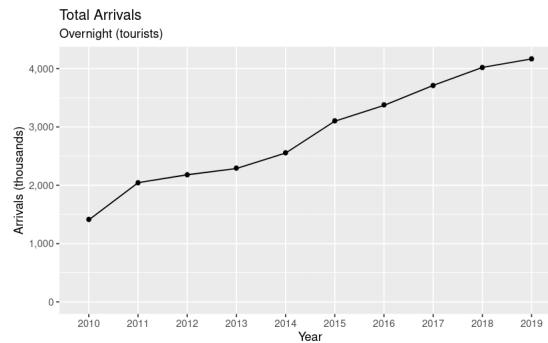
INBOUND TOURISM

Figure 2.1: UNWTO - Country Fact Sheets: Colombia - In bound tourism

Result:

	date	visitors
1	"2019-01-02T00:00:00"ˆˆ xsd:dateTime	"2744"ˆˆxsd:integer
2	"2019-01-03T00:00:00"ˆˆxsd:dateTime	"2380"ˆˆxsd:integer
3	"2019-01-04T00:00:00"ˆˆxsd:dateTime	"1736"ˆˆxsd:integer
4	"2019-01-05T00:00:00"ˆˆxsd:dateTime	"2118"ˆˆxsd:integer
5	"2019-01-06T00:00:00"ˆˆxsd:dateTime	"2304"ˆˆxsd:integer
6	"2019-01-07T00:00:00"ˆˆxsd:dateTime	"2904"ˆˆxsd:integer
7	"2019-01-08T00:00:00"ˆˆxsd:dateTime	"3334"ˆˆxsd:integer
8	"2019-01-09T00:00:00"ˆˆxsd:dateTime	"2413"ˆˆxsd:integer
9	"2019-01-10T00:00:00"ˆˆxsd:dateTime	"2332"ˆˆxsd:integer
10	"2019-01-11T00:00:00"ˆˆxsd:dateTime	"1965"ˆˆxsd:integer

Interpretation: From the expert's source, we can see 4,100,000 visitors in 2019. The execution of this query gives us a daily average of 2,423 reviews. Therefore, Booking reviews represent 21.57% of visitors to Colombia.

2.4 Test case 4: Impact on the destination of the offer of accommodation companies used by visitors

Listing 2.4: KPI-04 Number of tourism enterprises (accommodation) per 10000 population

```
PREFIX ott:    <http://tourdata.org/ontotoutra/ontotoutra.owl#>
PREFIX rdf:    <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX gn:     <http://www.geonames.org/ontology#>
PREFIX wgs84:  <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX xsd:    <http://www.w3.org/2001/XMLSchema#>

SELECT ((COUNT(?hotel) / xsd:integer(?population)*10000) AS ?hotels)
WHERE {
  ?hotel    rdf:type             ott:Hotel         ;
            ott:hasCityParent    ?city             .
  ?city     ott:hasStateParent   ?state            .
  ?state    ott:hasCountryParent ?country          .
  ?country  ott:countryName      ?countryName      .
  ?geo      gn:alternateName     ?alternateName    ;
            gn:population         ?population       .
  FILTER(?countryName = "Colombia")                .
  FILTER(CONTAINS(?alternateName, ?countryName)).
  FILTER(LANG(?alternateName) = "es")              .
}
GROUP BY ?population
```
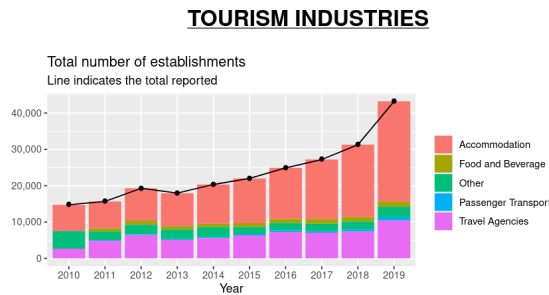
Expert: UNWTO

TOURISM INDUSTRIES

Figure 2.2: UNWTO - Country Fact Sheets: Colombia - Accommodation companies

Result: **"2.32755409332593602429"^^xsd:decimal**

Interpretation: From the expert's source, we can see around 28,000 accommodation establishments. Considering the population of 50 million inhabitants for Colombia in 2019, we would have a ratio of 5.6 establishments for every 10,000 inhabitants. After executing the query, we obtained a proportion of

2.33 establishments, which means that more than half of the accommodation establishments are registered with Booking.com.

2.5 Test case 5: Impact of visits on the destination

Listing 2.5: KPI-05 Ratio of number of reviews to local population

```
PREFIX gn:     <http://www.geonames.org/ontology#>
PREFIX wgs84: <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX ott:    <http://tourdata.org/ontotoutra/ontotoutra.owl#>
PREFIX rdf:    <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd:    <http://www.w3.org/2001/XMLSchema#>
PREFIX owl:    <http://www.w3.org/2002/07/owl#>

SELECT
  ?stateName
  (MAX(COALESCE(?reviews, 0)) AS ?rev)
  (MAX(COALESCE(?statePopulation, 0)) AS ?pop)
{
  {
    SELECT ?stateName (SUM(?hotelReviewNumber) AS ?reviews)
    WHERE {
      ?hotel    ott:hotelReviewNumber ?hotelReviewNumber ;
                ott:hasCityParent      ?city              .
      ?city     ott:hasStateParent     ?state             .
      ?state    ott:stateName          ?stateName         ;
                ott:hasCountryParent   ?country           .
      ?country  ott:countryName        ?countryName       .
      FILTER(?countryName = "Colombia")
    }
    GROUP BY ?stateName
    ORDER BY ?stateName
  }
  UNION
  {
    SELECT ?stateName (SUM(?population) AS ?statePopulation) {
      SELECT DISTINCT ?cityName ?stateName ?population
      WHERE {
        ?hotel    ott:hasCityParent      ?city             .
        ?city     ott:cityName           ?cityName         ;
                  ott:hasStateParent     ?state            .
        ?state    ott:stateName          ?stateName        ;
                  ott:hasCountryParent   ?country          .
        ?country  ott:countryName        ?countryName      .
        ?geo      gn:name                ?name             ;
                  gn:population          ?geopopulation    ;
                  gn:parentFeature       ?parent           .
        ?parent   gn:name                ?parentName       .
        BIND(REPLACE(REPLACE(REPLACE(REPLACE(REPLACE(
          ?name, "ú", "u", "i"),
          "ó", "o", "i"), "í", "i", "i"), "é", "e", "i"),
          "á", "a", "i") AS ?acc_name)
        BIND(REPLACE(REPLACE(REPLACE(REPLACE(REPLACE(REPLACE(REPLACE
          (?cityName, "ú", "u", "i"),
          "ó", "o", "i"), "í", "i", "i"), "é", "e", "i"),
          "á", "a", "i"), "city", "", "i"), "DC", "")
          AS ?acc_cityName)
        FILTER(CONTAINS(?acc_name, ?acc_cityName))
```

```
        BIND(REPLACE(REPLACE(REPLACE(REPLACE(REPLACE(
          ?parentName, "ú", "u", "i"),
          "ó", "o", "i"), "í", "i", "i"), "é", "e", "i"),
          "á", "a", "i") AS ?acc_parentName)
        BIND(REPLACE(REPLACE(REPLACE(REPLACE(REPLACE(
          ?stateName, "ú", "u", "i"),
          "ó", "o", "i"), "í", "i", "i"), "é", "e", "i"),
          "á", "a", "i") AS ?acc_stateName)
        FILTER((CONTAINS(?acc_parentName, ?acc_stateName)))
        FILTER(?countryName = "Colombia")
        BIND(xsd:integer(?geopopulation) AS ?population)
      }
    }
    GROUP BY ?stateName
    ORDER BY ?stateName
  }
}
GROUP BY ?stateName
ORDER BY ?stateName
```

Expert: MinCIT

Llegadas de pasajeros nacionales en vuelos regulares por principales aeropuertos
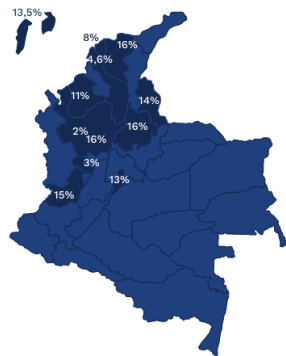
Aeropuerto	2018	2019	% Var
Bogotá - Eldorado	8.504.778	9.589.282	12,8%
Rionegro - Jose M. Córdova	2.967.374	3.445.580	16,1%
Cartagena - Rafael Núñez	2.133.115	2.230.225	4,6%
Cali - Alfonso Bonilla Aragón	1.797.887	2.073.893	15,4%
Barranquilla-E. Cortissoz	1.067.791	1.157.785	8,4%
Santa Marta - Simón Bolívar	977.562	1.134.924	16,1%
San Andres - Gustavo Rojas Pinilla	944.013	1.071.036	13,5%
Bucaramanga - Palonegro	740.434	864.320	16,7%
Pereira - Matecañas	746.993	773.019	3,5%
Medellín - Olaya Herrera	520.261	530.615	2,0%
Montería - Los Garzones	456.723	505.604	10,7%
Otros	2.492.099	2.884.664	15,8%
Total General	23.349.030	26.260.947	12,5%

Fuente: Aeronáutica Civil: Boletín origen-destino, diciembre 2019. Cálculos OEE – MinCIT.

Figure 2.3: MinCIT - Colombia - Local Arrivals

Result:

stateName	reviews	population
Amazonas	5,305	44,815
Antioquia	146,842	3,830,053
Arauca	330	75,557
Atlántico	33,620	1,941,838

Bogotá	174,776	6,840,116
Bolívar	164,167	1,223,076
Boyacá	39,191	864,913
Caldas	11,254	720,124
Caquetá	419	143,871
Casanare	2,313	239,953
Cauca	5,414	448,882
Cesar	5,206	503,654
Choco	3,799	151,909
Cundinamarca	34,142	1,999,812
Córdoba	7,790	836,259
Guainía	102	17,866
Guaviare	146	53,994
Huila	14,366	722,757
La Guajira	21,765	515,117
Magdalena	98,522	656,825
Meta	9,409	598,295
Nariño	7,928	701,453
Norte de Santander	9,463	839,131
Putumayo	744	214,182
Quindío	44,912	208,314
Risaralda	20,042	866,643
San Andrés y Providencia	48,310	65,627
Santander	39,586	1,614,902
Sucre	8,780	483,695
Tolima	16,370	850,170
Valle del Cauca	44,140	3,774,893
Vaupés	26	28,382
Vichada	0	28,718

Interpretation: according to the expert, the three airports with the highest national passengers are Bogotá, Ríonegro (Antioquia), and Cartagena (Bolivar). This Top-3 coincides with the query of OntoTouTra, Bogotá 174,776 reviews, Bolivar 164,167 reviews, and Antioquia 146,842 reviews.

2.6 Test case 6: Influence of accommodation companies in the destination

Listing 2.6: KPI-06 Population rate with hotel influence

```
PREFIX gn:     <http://www.geonames.org/ontology#>
PREFIX wgs84:  <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX ott:    <http://tourdata.org/ontotoutra/ontotoutra.owl#>
PREFIX rdf:    <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd:    <http://www.w3.org/2001/XMLSchema#>
PREFIX owl:    <http://www.w3.org/2002/07/owl#>

SELECT ?stateName (?citiesPop / ?statePop * 100 AS ?populationRate){
  SELECT
    ?stateName
    (SUM(?cityPopulation) AS ?citiesPop)
    (MAX(?statePopulation) as ?statePop) {
      SELECT DISTINCT
        ?cityName ?stateName ?cityPopulation ?statePopulation
      WHERE {
        ?hotel     ott:hasCityParent     ?city            .
        ?city      ott:cityName          ?cityName        ;
                   ott:hasStateParent    ?state           .
        ?state     ott:stateName         ?stateName       ;
                   ott:hasCountryParent  ?country         .
        ?country   ott:countryName       ?countryName     .
        ?geo       gn:name               ?name            ;
                   gn:population         ?population      ;
                   gn:parentFeature      ?parent          .
        ?parent    gn:name               ?parentName      ;
                   gn:population         ?parentPopulation .
        BIND(REPLACE(REPLACE(REPLACE(REPLACE(REPLACE(
          ?name, "ú", "u", "i"),
          "ó", "o", "i"), "í", "i", "i"), "é", "e", "i"),
          "á", "a", "i")
          AS ?acc_name)
        BIND(REPLACE(REPLACE(REPLACE(REPLACE(REPLACE(REPLACE(REPLACE
          (?cityName, "ú", "u", "i"),
          "ó", "o", "i"), "í", "i", "i"), "é", "e", "i"),
          "á", "a", "i"), "city", "", "i"), "DC", "")
          AS ?acc_cityName)
        FILTER(CONTAINS(?acc_name, ?acc_cityName))
        BIND(REPLACE(REPLACE(REPLACE(REPLACE(REPLACE(
          ?parentName, "ú", "u", "i"),
          "ó", "o", "i"), "í", "i", "i"), "é", "e", "i"),
          "á", "a", "i") AS ?acc_parentName)
        BIND(REPLACE(REPLACE(REPLACE(REPLACE(REPLACE(
          ?stateName, "ú", "u", "i"),
          "ó", "o", "i"), "í", "i", "i"), "é", "e", "i"),
          "á", "a", "i") AS ?acc_stateName)
        FILTER((CONTAINS(?acc_parentName, ?acc_stateName)))
        FILTER(?countryName = "Colombia")
        BIND(xsd:integer(?population) AS ?cityPopulation)
        BIND(xsd:integer(?parentPopulation) AS ?statePopulation)
      }
    }
```

```
    GROUP BY ?stateName
    ORDER BY ?stateName
}
```

Expert: MinCIT

Figure 2.4: MinCIT - Colombia - Local Tourism Industry

Result:

stateName	populationRate
Amazonas	66.17104214038921536780566
Antioquia	61.87496091199110587080466
Arauca	32.55111624260074617220555
Atlántico	89.64442080810431012355553
Bogotá	100.0
Bolívar	59.688943786074063373714
Boyacá	73.31155386991749454915955
Caldas	75.438507752338088651237777
Caquetá	34.227536476684184356837555
Casanare	85.128981252941395550409222
Cauca	35.374648229187107003736222
Cesar	55.75840908512209405953222
Choco	33.457921282734621060282444
Cundinamarca	77.232439589030287751924999
Córdoba	56.968672292461011397699
Guainía	50.71246097076355378938444
Guaviare	56.508042825297485112662333

Huila	71.4597723196541884759812
La Guajira	78.6314052011884238711807
Magdalena	57.1193399175766598806696
Meta	85.0181570237803383182152
Nariño	45.4911164780317985727219
Norte de Santander	67.455616069454772001045
Putumayo	69.0615608837527246462796
Quindío	100.0
Risaralda	96.5609258514399298502856
San Andrés y Providencia	93.0166964311024180060663
Santander	82.7718410921708110526722
Sucre	62.6539811660470719291201
Tolima	65.3637696635714714701518
Valle del Cauca	90.7115471262848663618832
Vaupés	72.2574403625346877466331
Vichada	51.3996277205040091638303

Note: In the results of Listing 2.6, some departments appear with values of 100. It means that all the municipalities of that department have hotel influence.

Interpretation: In the expert's data source, the three departments with the most significant hotel influence according to their population are: San Andrés, Bolivar, and Bogotá. After executing the query, we obtained similar results except for Bolivar.

2.7 Test case 7: Arrival of foreign tourists (FTA)

Listing 2.7: KPI-07 Foreign Tourist Arrivals (FTAs) - Top 10

```
PREFIX ott:    <http://tourdata.org/ontotoutra/ontotoutra.owl#>
PREFIX rdf:    <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd:    <http://www.w3.org/2001/XMLSchema#>
PREFIX owl:    <http://www.w3.org/2002/07/owl#>

SELECT ?countryName (COUNT(?review) AS ?visitors)
WHERE {
   ?review   ott:hotelReviewID      ?hotelReviewID    ;
             ott:hasCountryParent ?country            .
   ?country  ott:countryName         ?countryName     .
   FILTER(?countryName != "Colombia")
}
GROUP BY ?countryName
ORDER BY DESC(?visitors)
LIMIT 10
```

Expert: MinCIT
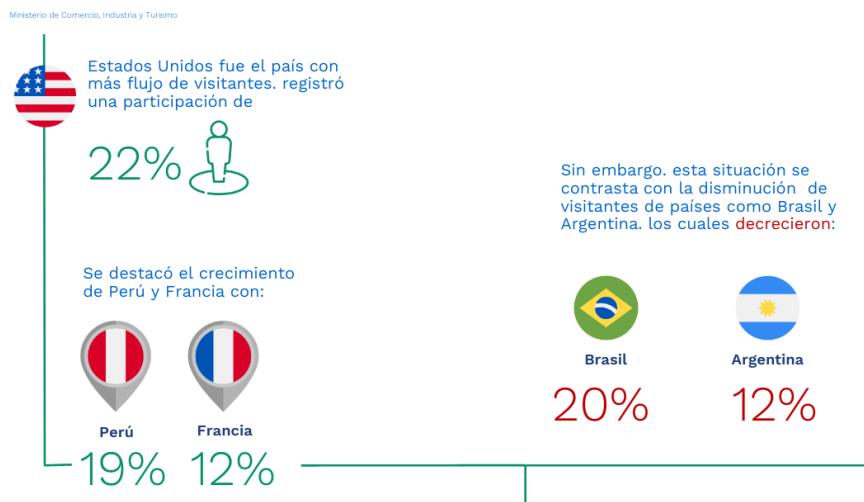
Figure 2.5: MinCIT - Colombia - FTA

Result:

countryName	visitors
USA	42,422
France	40,183
Argentina	39,753
Germany	34,225

Spain	31,420
Brazil	25,724
The Netherlands	24,447
Chile	18,878
United Kingdom	18,476
Italy	14,470

Interpretation: The provenance of foreign tourists is very similar to that reported by the expert with the query results in OntoTouTra.

2.8 Test case 8: Inbound and local tourism

Listing 2.8: KPI-08 Inbound and domestic tourism

```
PREFIX ott:   <http://tourdata.org/ontotoutra/ontotoutra.owl#>
PREFIX rdf:   <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd:   <http://www.w3.org/2001/XMLSchema#>
PREFIX owl:   <http://www.w3.org/2002/07/owl#>

SELECT
  ?stateName
  (MAX(COALESCE(?domestic, 0)) AS ?national)
  (MAX(COALESCE(?foreign, 0)) AS ?international)
{
  {
    SELECT ?stateName (COUNT(?review) AS ?domestic)
    WHERE {
      ?review   ott:hotelReviewID     ?hotelReviewID    ;
                ott:hasHotel           ?hotel            ;
                ott:hasCountryParent   ?country          .
      ?hotel    ott:hasCityParent      ?city             .
      ?city     ott:hasStateParent     ?state            .
      ?state    ott:stateName          ?stateName        .
      ?country  ott:countryName        ?countryName      .
      FILTER(?countryName = "Colombia")
    }
    GROUP BY ?stateName
    ORDER BY ?stateName
  }
  UNION
  {
    SELECT ?stateName (COUNT(?review) AS ?foreign)
    WHERE {
      ?review   ott:hotelReviewID     ?hotelReviewID    ;
                ott:hasHotel           ?hotel            ;
                ott:hasCountryParent   ?country          .
      ?hotel    ott:hasCityParent      ?city             .
      ?city     ott:hasStateParent     ?state            .
      ?state    ott:stateName          ?stateName        .
      ?country  ott:countryName        ?countryName      .
      FILTER(?countryName != "Colombia")
    }
    GROUP BY ?stateName
    ORDER BY ?stateName
  }
}
GROUP BY ?stateName
ORDER BY ?stateName
```

Expert: UNWTO See Figure 2.1.

Result:

stateName	national	international
Amazonas	2,137	3,156

Antioquia	78,038	68,277
Arauca	287	42
Atlántico	25,317	8,213
Bogotá	83,918	97,460
Bolívar	62,346	101,641
Boyacá	31,753	7,445
Caldas	7,951	3,294
Caquetá	362	55
Casanare	2,145	165
Cauca	3,003	2,408
Cesar	4,740	466
Choco	2,043	1,748
Cundinamarca	30,299	4,529
Córdoba	7,108	670
Guainía	87	14
Guaviare	104	42
Huila	8,824	5,514
La Guajira	10,843	10,892
Magdalena	45,086	50,901
Meta	8,712	696
Nariño	4,216	3,703
Norte de Santander	4,493	5,078
Putumayo	472	270
Quindío	24,645	23,407
Risaralda	15,842	5,044
San Andrés y Providencia	20,610	27,601
Santander	29,755	9,792
Sucre	7,525	2,280
Tolima	15,709	1,222
Valle del Cauca	30,586	13,313
Vaupés	22	4
Total	468,978	459,322

Interpretation: According to the expert's data, inbound tourism was around 4,100,000 visitors for 2019. When executing our query, we obtained 459,322 visits, which is equivalent to 11

2.9 Test case 9: Seasonality patterns in the destination

Listing 2.9: KPI-09 Seasonality Patterns

```
PREFIX ott:    <http://tourdata.org/ontotoutra/ontotoutra.owl#>
PREFIX rdf:    <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd:    <http://www.w3.org/2001/XMLSchema#>
PREFIX owl:    <http://www.w3.org/2002/07/owl#>

SELECT ?month (COUNT(?review) AS ?visitors) {
  ?review   ott:hotelReviewID   ?reviewID   ;
            ott:hotelReviewDate ?reviewDate .
  FILTER(year(?reviewDate) = 2019)
}
GROUP BY (month(?reviewDate) AS ?month)
ORDER BY ?month
```

Expert: UNWTO
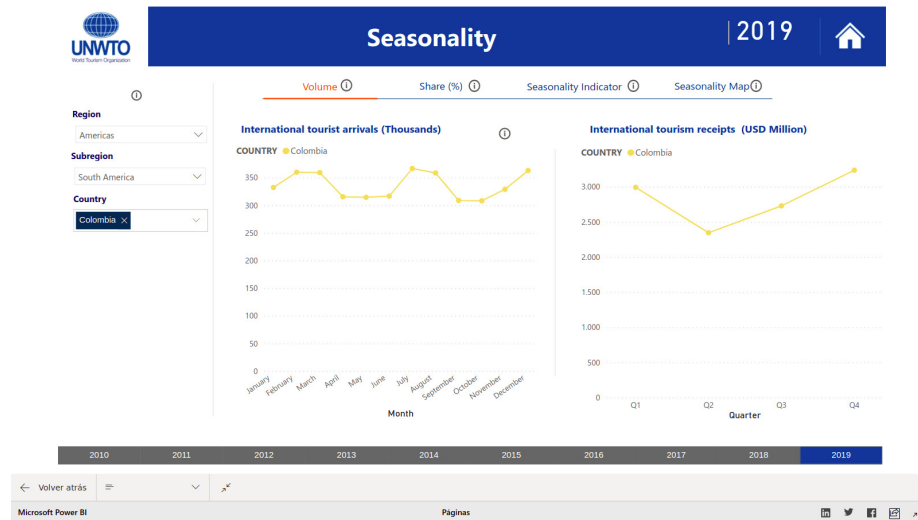
Figure 2.6: UNWTO - Colombia - Tourist Seasonality

Result:

month	reviews
01	59,177
02	41,501
03	47,009
04	46,516
05	37,277

06	41,637
07	54,123
08	60,274
09	48,215
10	47,720
11	49,521
12	41,421

Interpretation - The expert's data source matches the results of the season-ality query. Two peaks are observed per year, one between January and March and the other between July and August.

2.10 Test case 10: Portfolio of tourist experiences used

Listing 2.10: KPI-10 Tourist experiences - Top 10

```
PREFIX ott:    <http://tourdata.org/ontotoutra/ontotoutra.owl#>
PREFIX rdf:    <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd:    <http://www.w3.org/2001/XMLSchema#>
PREFIX owl:    <http://www.w3.org/2002/07/owl#>

SELECT ?activity (COUNT(?hotel) AS ?hotels) {
  ?hotel      ott:hasService            ?service       .
  ?service    ott:hasServiceCategory    ?category      ;
              ott:serviceName           ?activity      .
  ?category  ott:serviceCategoryName ?categoryName .
  FILTER(STR(?categoryName) = "Activities")
  FILTER(LANG(?categoryName) = "en")
}
GROUP BY ?activity
```

Result:

activity	hotels
Beach	1,621
Walking tours	1,267
Game room	1,098
Bike tours Additional charge	1,029
Tour or class about local culture	966
Bicycle rental (for a fee)	941
Cycling Outside the accommodation	756
Trekking	752
Walking tours	698
Hiking Outside the accommodation	683

Interpretation: regarding tourist experiences, we did not find an official source from Colombia. However, our query in OntoTouTra highlighted experiences such as beach tourism, tours, and game rooms as the three most offered experiences by tourist providers in Colombia.

27

Chapter 3

Big Data Analytics Lifecycle for building TTS ontology

We adapt the Big Data life cycle's Erl [4] methodology to construct our ontology in this type of environment. Below we will show some source code listings that implement the essential phases of this life cycle to illustrate this implementation with an actual use case.

3.1 Definition of the ontology purpose

Listing 3.1: OntoTouTra preamble

```xml
<?xml version="1.0"?>
<rdf:RDF xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:xml="http://www.w3.org/XML/1998/namespace"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:ontotoutra="http://tourdata.org/ontotoutra/ontotoutra.owl#">
<owl:Ontology rdf:about="http://tourdata.org/ontotoutra/ontotoutra.owl">
<owl:versionIRI rdf:resource="http://tourdata.org/ontotoutra/ontotoutra.owl/
1.0.0"/>
  <rdfs:comment xml:lang="en">
    tourist traceability is the analysis of the set of actions,
    procedures, and technical measures that allows us to identify
    and record the time−space relationship of the touring,
    from the beginning to the end of the chain of tourist products.
  </rdfs:comment>
</owl:Ontology>
```

3.2 Data Validation & Cleansing

In Listing 3.2, we see the validation and cleaning of the general data of a hotel, in this case, the integer or float data type for the hotel or destination: id, score, and the number of reviews made.

Listing 3.2: Data type validation for general destination or hotel fields

```python
hotelsDF['hotelID'] = hotelsDF['hotelID'].apply(lambda x : int(x))
hotelsDF['reviewScore'] = hotelsDF['reviewScore'].apply(
  lambda  x: 0 if x is None else float(x.replace(',', '.'))
)
hotelsDF['reviewNumber'] = hotelsDF['reviewNumber'].apply(
  lambda x: 0 if x is None else int(x.split()[0].replace('.', ''))
)
```

3.3 Data Aggregation & Representation

In Listing 3.3, we see an example of unification of datasets of reviews of some locations, using Apache Spark, which is an open-source cluster computing framework widely used in Big Data environments. We want to detect the 20 most used words in these reviews and how often these were used through clustering. The results are seen in Listing 3.4.

Listing 3.3: Word counts of reviews using Apache Spark

```python
from pyspark import SparkContext, SparkConf

if __name__ == "__main__":
  conf = SparkConf().setAppName("word count").setMaster("local[3]")
  sc = SparkContext(conf = conf)

  lines = sc.textFile("locationOnlyReviews.txt")
  words = lines.flatMap(lambda line: line.split(" "))
  wordCounts = words.countByValue()
  result = sorted(
    wordCounts.items(), key=lambda x:x[1], reverse=True
  )

  index = 0
  for word, count in result:
    if len(word) > 0:
      print("{} : {}".format(word, count))
      index += 1
    if index >= 20:
      break
```

Listing 3.4: Results of word counts of reviews using Apache Spark

```
$ spark−submit WordCount.py
the    : 2149
and    : 1913
to     : 1684
```

```
a       :  1314
is      :  1120
of      :  871
in      :  646
for     :  505
with    :  451
The     :  445
are     :  428
you     :  402
very    :  365
but     :  352
was     :  344
I       :  334
place   :  307
great   :  281
nice    :  280
town    :  279
```

3.4   Data analysis

In Listings 3.5 and 3.6, we can see the Python code snippet to convert the data
from Web Scraping into triples of the OntoTouTra ontology (subject, predicate,
and object). In the first four lines of the result, we can see the destination data
from Web Scraping and stored in a dataset. Next, the snippet code displays
the RDF representation of the destination data, as the destination belongs to
a state. Through a SPARQL query, we obtained the state's name to establish
the "hasStateParent" relationship. Finally, the execution of this code displays
the triples in turtle format of the same RDF listing shown. The example only
shows the data for one tourist destination in the dataset.

Listing 3.5: Generating ontology triples with RDFLib

```python
import rdflib
from rdflib.namespace import FOAF, DCTERMS, XSD, RDF, SDO
from rdflib import URIRef, BNode, Literal, Namespace

g = rdflib.Graph()
onto_filename = os.path.join(path, 'ontotoutra.owl')

format_ = rdflib.util.guess_format(onto_filename)
g.parse(onto_filename, format=format_)

qres = g.query('''
prefix xsd:  <http://www.w3.org/2001/XMLSchema#>
PREFIX my:   <http://tourdata.org/ontotoutra/ontotoutra.owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?stateID ?stateName
WHERE {
        ?state my:stateID ?stateID ;
        my:stateName ?stateName .
      }
''')
```

```
states = {}
for stateID, stateName in qres:
    states[stateID.value] = stateName

ott = Namespace('http://tourdata.org/ontotoutra/ontotoutra.owl#')
h = rdflib.Graph()
h.bind('ott', ott)

for index, row in cities_df.iterrows():
    cityID    = row['cityID']
    cityName  = row['cityName']
    stateID   = row['stateID']
    stateName = ott + states[stateID]

    city = ott[cityName.replace(" ", "")]
    h.add((city, RDF.type, ott.City))
    h.add((city, ott.hasStateParent, Literal(stateName)))
    h.add((city, ott.cityID, Literal(cityID, datatype=XSD.integer)))
    h.add((city, ott.cityName, Literal(cityName)))
    h.add((city, ott.stateID, Literal(stateID, datatype=XSD.integer)))

print(row)
print(h.serialize(format='xml').decode('u8'))
print(h.serialize(format='ttl').decode('u8'))
```

Listing 3.6: Results of generating ontology triples with RDFLib

```
cityID             1
cityName      Leticia
stateID         5131
Name: 0, dtype: object

<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
xmlns:ott="http://tourdata.org/ontotoutra/ontotoutra.owl#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
>
  <rdf:Description rdf:about="http://tourdata.org/ontotoutra
    /ontotoutra.owl#Leticia">
    <ott:cityName>Leticia</ott:cityName>
    <rdf:type rdf:resource="http://tourdata.org
      /ontotoutra/ontotoutra.owl#City"/>
    <ott:cityID rdf:datatype="http://www.w3.org/2001
      /XMLSchema#integer">1</ott:cityID>
    <ott:hasStateParent>
       http://tourdata.org/ontotoutra/ontotoutra.owl#Amazonas
    </ott:hasStateParent>
    <ott:stateID
       rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">5131
    </ott:stateID>
  </rdf:Description>
</rdf:RDF>

@prefix ott: <http://tourdata.org/ontotoutra/ontotoutra.owl#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
```

```
ott:Leticia a ott:City ;
ott:cityID 1 ;
ott:cityName "Leticia" ;
ott:hasStateParent "http://tourdata.org/ontotoutra
  /ontotoutra.owl#Amazonas" ;
ott:stateID 5131 .
```

Chapter 4

Screenshots of SPARQL queries

Figures 4.1, 4.2, 4.3, and 4.4 show the execution of the SPARQL queries on different endpoints: Apache Fuseki, Apache Jena, Protégé, and Open Link Virtuoso. It demonstrates the interoperability of the ontology.

Figure 4.1: SPARQL query in OntoTouTra using Apache Fuseki

```
(base) [jf@fedora 06 ontology]$ cat q1.rq
PREFIX ott: <http://tourdata.org/ontotoutra/ontotoutra.owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT ?countryName ?p ?o
WHERE {
    ?country rdf:type          ott:Country  ;
             ott:countryName ?countryName .
    ?country ?p                ?o           .
    FILTER (?countryName = 'France')
}
(base) [jf@fedora 06 ontology]$ sparql --data=ontotoutra.owl --query=q1.rq

---------------------------------------------------------------------------
| countryName | p                | o                                      |
===========================================================================
| "France"    | ott:alpha2Code   | "FR"                                   |
| "France"    | ott:alpha3Code   | "FRA"                                  |
| "France"    | rdf: type        | ott:Country                            |
| "France"    | rdf: type        | <http://www.w3.org/2002/07/owl#NamedIndividual> |
| "France"    | ott:countryName  | "France"                               |
| "France"    | ott:countryID    | 250                                    |
---------------------------------------------------------------------------
```

Figure 4.2: SPARQL query in OntoTouTra using Apache Jena
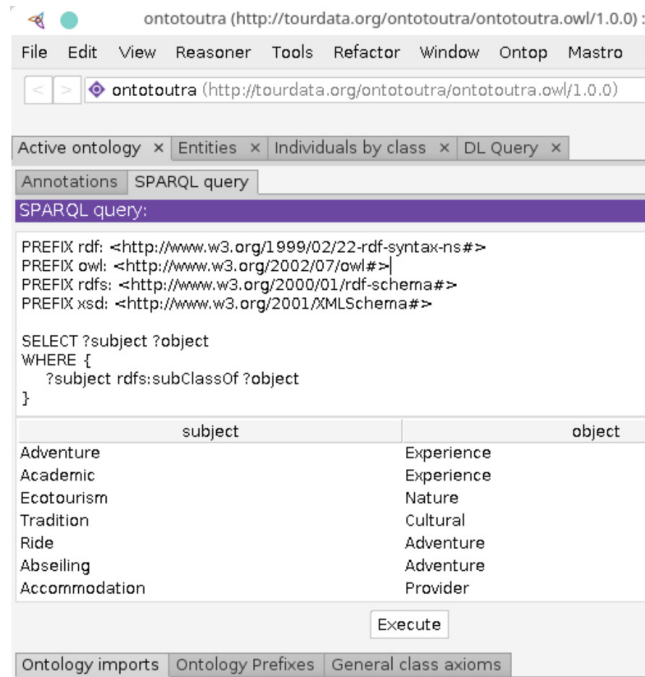
Figure 4.3: SPARQL query in OntoTouTra using Protégé
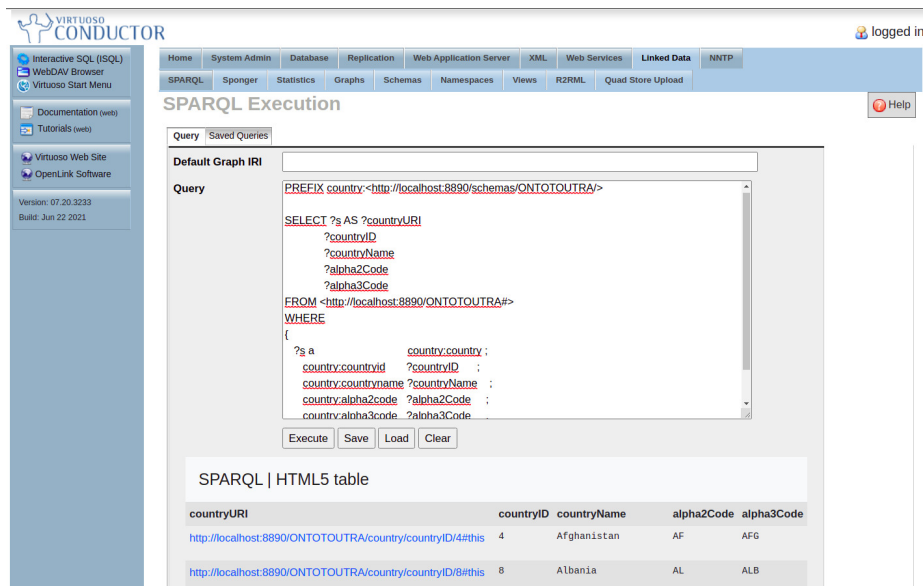
Figure 4.4: SPARQL query in OntoTouTra using OpenLink Virtuoso

```
(base) [jf@fedora 10_RESTFul]$ s-query --service http://localhost:3030/
ds/query 'PREFIX ott: <http://tourdata.org/ontotoutra/ontotoutra.owl#>

SELECT ?cityID ?cityName ?stateName
WHERE {
   ?state ott:stateName    ?stateName ;
          ott:stateCapital ?cityName  .
   FILTER(?stateName = "Boyaca")
}'
{ "head": {
    "vars": [ "cityID" , "cityName" , "stateName" ]
  } ,
  "results": {
    "bindings": [
      {
        "cityName": { "type": "literal" , "value": "Tunja" } ,
        "stateName": { "type": "literal" , "value": "Boyaca" }
      }
    ]
  }
}
```

Figure 4.5: REST API in OntoTouTra using Fuseki SOH

Figure 4.6: REST API in OntoTouTra using OBA OpenAPI

In Figures 4.7 and 4.8, we see the documentation generated by OntoTouTra from two different systems: Protégé and OBA, respectively.

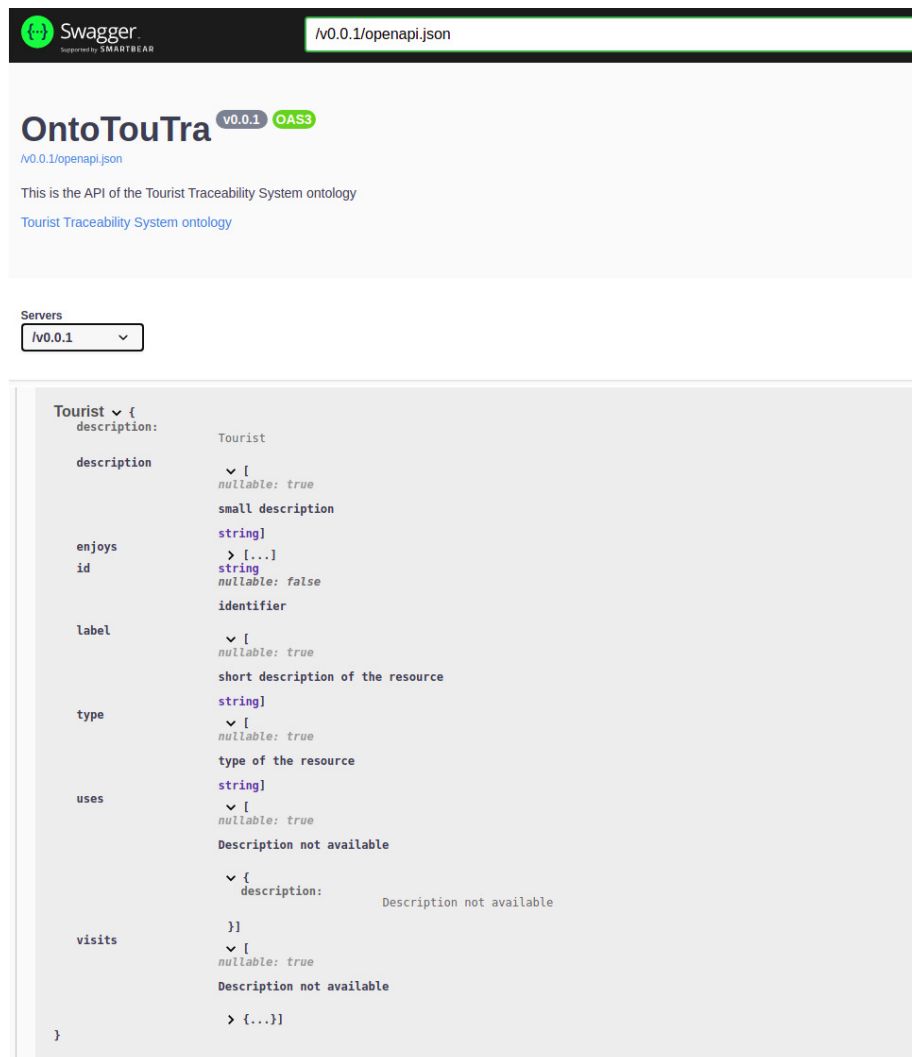Figure 4.7: OntoTouTra documentation generated by Protégé

Figure 4.8: OntoTouTra documentation generate by OBA

Disclaimer

We recommend that whoever uses the software that accompanies this paper use it responsibly. To avoid legal problems, check the Web Site's rules, the data provider device, or the application installed on the user's device.

Abbreviations

Bibliography

[1] J. Brank, M. Grobelnik, and D. Mladenić, "A survey of ontology evaluation techniques," 01 2009.

[2] M. Zarate, G. Braun, P. Fillottrani, C. Delrieux, and M. Lewis, "BiGe-Onto: An ontology-based system for managing biodiversity and biogeography data," *Applied ontology*, pp. 1–27, 05 2020.

[3] Office for National Statistics, *Measuring Tourism Locally*, S. White, Ed. ONS, 2010.

[4] T. Erl, W. Khattak, and P. Buhler, *Big Data Fundamentals: Concepts, Drivers & Techniques.* ServiceTech Press, 2016.