TIN2009-11005
DAMASK

Data-Mining Algorithms with Semantic Knowledge

PROYECTO DE INVESTIGACIÓN
PROGRAMA NACIONAL DE INVESTIGACIÓN FUNDAMENTAL,
PLAN NACIONAL DE I+D+i 2008-2011
ÁREA TEMÁTICA DE GESTIÓN: Tecnologías informáticas

Internal project report T3.1

Damask Ontology

Authored by
Carlos Vicient, Universitat Rovira i Virgili

Co-authored by
Antonio Moreno, Universitat Rovira i Virgili

Document information

project name:	DAMASK	
Project reference:	TIN2009-11005	
type of document:	Deliverable	
file name:	ProjRepT3.1	
version:	final	
authored by:	C.Vicient	14/10/2011
co-authored by	A.Moreno	
released by:	Co-ordinator	
approved by:	Co-ordinator	A. Moreno

Document history

version	date	reason of modification
0.1	14.October.2011	A preliminary release of the document.
1.0	20.Nov.2001	Final version
	. .200	
	. .200	
	. .200	
	. .200	
	. .200	

Table of Contents

1 Introduction

The first task of DAMASK, called T1 - *Semantic integration of the information available in heterogeneous Web resources*, includes two preliminary subtasks, tasks 1.1 and 1.2 (see Figure 1). The former task discusses all related works in information extraction from the Web distinguishing algorithms to extract structured, semi-structured and non-structured resources. Deliverable D1 was the result of the task 1.1., and its aim was to make a state-of-the-art on Information Extraction techniques applied to Web resources. An internal report, which was the output of task 1.2, lists the types of data that can be extracted using the previous algorithms. As stated in that document, DAMASK takes into account three data types: Measurement, Nominal and Semantic. Finally, Deliverable D2 was the result of task 1.3. (see Figure 1), and its aim was to present the proposed methodology for Ontology-based feature extraction. The main goal of that task was to design and implement a novel method that was able to extract relevant features from a range of textual documents going from plain textual data to semi-structured resources. The designed methodology was able to take profit from pre-processed input when it is available in order to complement its own learning algorithms[1, 2]. The key point of the work was to complement the syntactical parsing and other natural language processing techniques with the knowledge contained in an input ontology (which ideally, should model the knowledge domain in which the posterior data analysis will be focused –e.g. touristic points of interest).

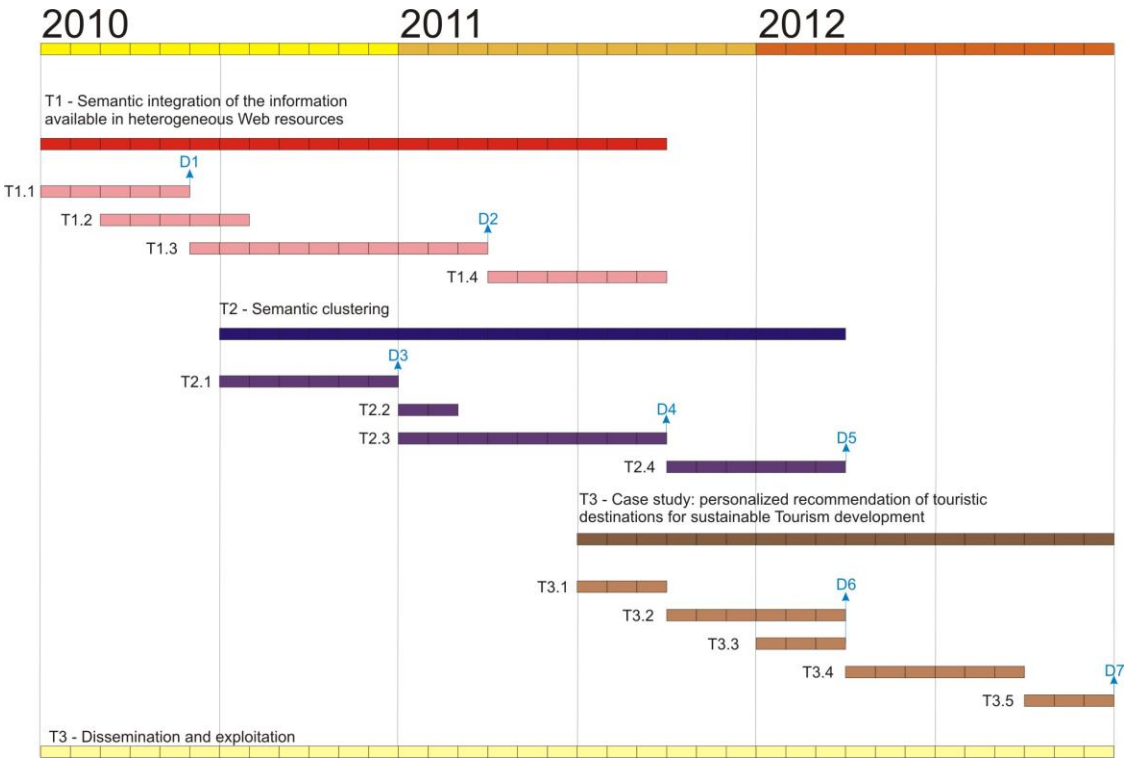


Figure 1: Tasks of DAMASK

Task T2 was focused on the goal O2 of the project: design of a clustering method based on ontologies. The inputs of this task were (1) a data matrix object × attribute (e.g. touristic destinations) and (2) a domain ontology. Based on those inputs, a method for automatically building clusters was needed. During the clustering, contextual knowledge provided by the domain ontology was used [3]. Finally, an automatic interpretation process of the clusters was required, in order to obtain a semantic description of the clusters that can help the user in his/her decision making tasks.

In task T3 the goal is to evaluate the deployment of the methods designed in the previous tasks in a particular case study: a personalized recommendation system of touristic destinations. A Web application will be designed to offer this kind of recommendations to any user. The tool will be focused on searching touristic destinations in the different types of touristic resources available in Internet using the tools developed in task T1. The clustering methods defined in T2 will then be applied to obtain a classification of touristic destinations based on the domain knowledge and the user preferences, in order to be able to recommend the set of places that match better with the user's interests. In task T3, the experience and touristic knowledge of the EPOs of the project is very valuable (EPO=organism or industry interested in the results of the project). They are particularly very helpful in the construction of the domain ontology and the evaluation of the recommendations provided by the system. Particularly, this document (Internal project report T3.1) is the result of the design and creation of the aforementioned ontology (Task 3.1). As this work is focused on the touristic domain, it has been implemented in order to model the most important aspects which represent a touristic destination such as monuments, landmarks, etc. To do so, three existent ontologies modelled in this domain were studied and merged.

2 Ontology

This section states some ontologies related in the area of tourism (section 2.1) and presents the proposed ontology for DAMASK (section 2.2).

2.1 Background ontologies

TourismOWL.owl ontology

This ontology models touristic points of interest for different kinds of tourist profiles. It was designed in a final year project [4] based on information extracted through Wikipedia articles. It consists of 315 classes and a depth of 5 hierarchical levels. Its main classes represent concepts related with administrative divisions (borough, city, country, village, etc.), buildings (commercial buildings, cultural buildings, religious buildings, sport buildings, etc.), festivals (art festivals, music festivals, carnival, etc.), landmarks (commemorate landmarks, geographical landmarks, memorial landmarks, etc.), museums (archaeology museum, history museum, science museum, etc.) and sports (football, basketball, hockey, formula one, etc.). See Annex I.

Space.owl ontology

This ontology was found by looking up ontologies using the Web search engine SWOOGLE[1] that is specialized in ontologies. The Space.owl[2] ontology consists of 188 classes and a depth of 6 hierarchical levels. It contains concepts related with three main topics: geographical features (i.e., archipelago, beach, river, forest, etc.), geopolitical entities (i.e., country, capital, city, district, street, etc.) and places, which includes business places (factory, convention centre, etc.), private places (residential structure, home, etc.) and public places (educational and medical structures, entertainment places, shopping facilities, transportation connections, etc.). See Annex II.

PCTTO.owl ontology

This ontology was created with the collaboration of the project EPOs: the Scientific and Technological Park for Tourism and Leisure of Vila-Seca (Tarragona), and the Tourism Observatory for Costa Daurada. It is focused on tourist activities. The ontology represents up to 203 connected concepts in 5 hierarchy levels. It is structured around eight main concepts, that constitute the first level of the hierarchy: "Events", "Nature", "Culture", "Leisure", "Sports", "Towns", "Routes" and "ViewPoints". The last three classes are considered transversal concepts, since they share children nodes with other main classes, e.g., "Routes" and "Nature" are both superclasses of the "NatureRoutes" class. The rest of the concepts in the ontology are connected via is-a (subclass) relationships with these main classes. The ontology is not a pure taxonomy, as it contains multi-inheritance between concepts, e.g., EthnographicMuseum is a subclass of both Museum and TraditionalCulturalElement. This ontology was developed using the Thesaurus of the World Tourism Organization as a reference guide to represent the touristic and leisure activities in the "Costa Daurada and Terres de l'Ebre" region. See Annex III

2.2 Damask Ontology

This ontology[3] is the result of merging and combining the aforementioned ontologies. It consists of 538 classes connected in 9 hierarchy levels. It is structured around 4 main concepts that constitute the first level of the hierarchy: "geopolitical division", "activity", "point of interest" and "geographical feature". The Damask Ontology is not a pure taxonomy, as it contains multi-inheritance between concepts. Annex IV depicts the taxonomy of those main concepts from the first level. The whole ontology is available at the DAMASK project Web page[4] (section "ontologies") in "png", "jpg" and "owl" formats.

[1] http://swoogle.umbc.edu/

[2] http://deim.urv.cat/~itaka/CMS2/images/ontologies/space.owl

[3] http://deim.urv.cat/~itaka/CMS2/images/ontologies/damaskontology.owl

[4] http://deim.urv.cat/~itaka/CMS2/index.php?option=com_content&task=view&id=29&Itemid=51

3 References

[1] C. Vicient, D. Sánchez, and A. Moreno, "Ontology-Based Feature Extraction," in *Proceedings of the IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, NLPOE*, Lyon, France, 2011, pp. 189-192.

[2] C. Vicient, D. Sánchez, and A. Moreno, "A methodology to discover semantic features from textual resources," in *Proceedings of the 6th International Workshop on Semantic media adaptation and personalization (SMAP 2011)*, Vigo, Spain, 2011.

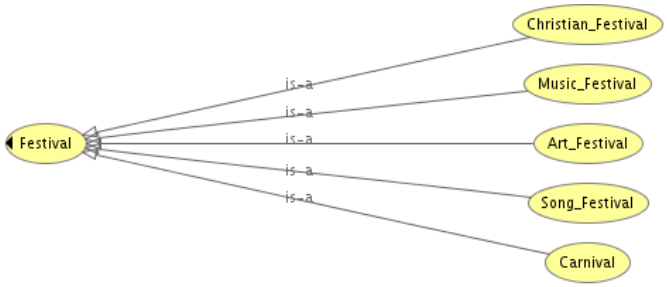
[3] M.Batet, "Ontology-based semantic clustering". PhD Thesis, Univ. Rovira i Virgili, Tarragona, February 2011.

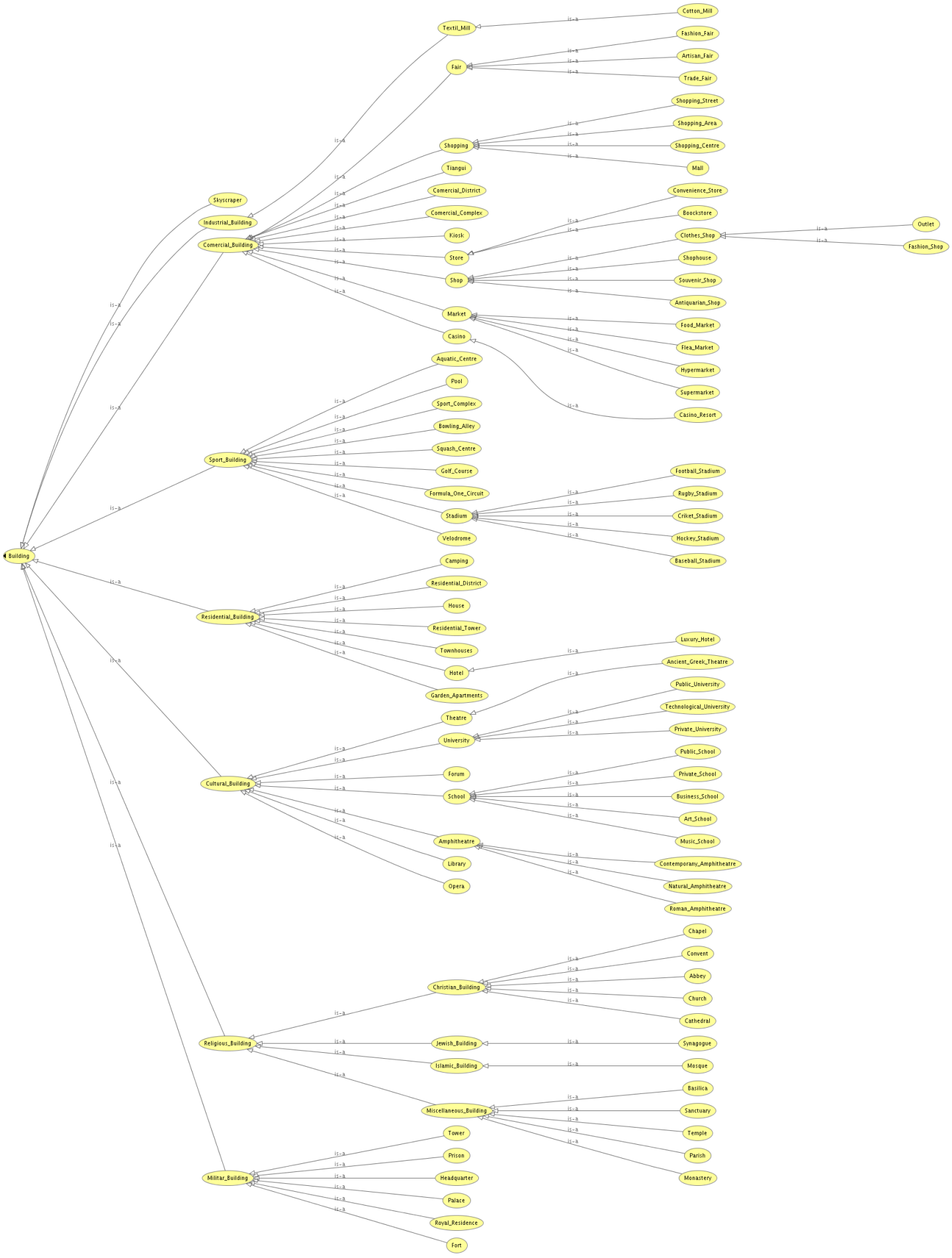
[4] C. Vicient, "Extracció basada en ontologies d'informació de destinacions turístiques a partir de la Wikipedia," . Computer Science Final Year Project, Universitat Rovira i Virgili, Tarragona, 2009.

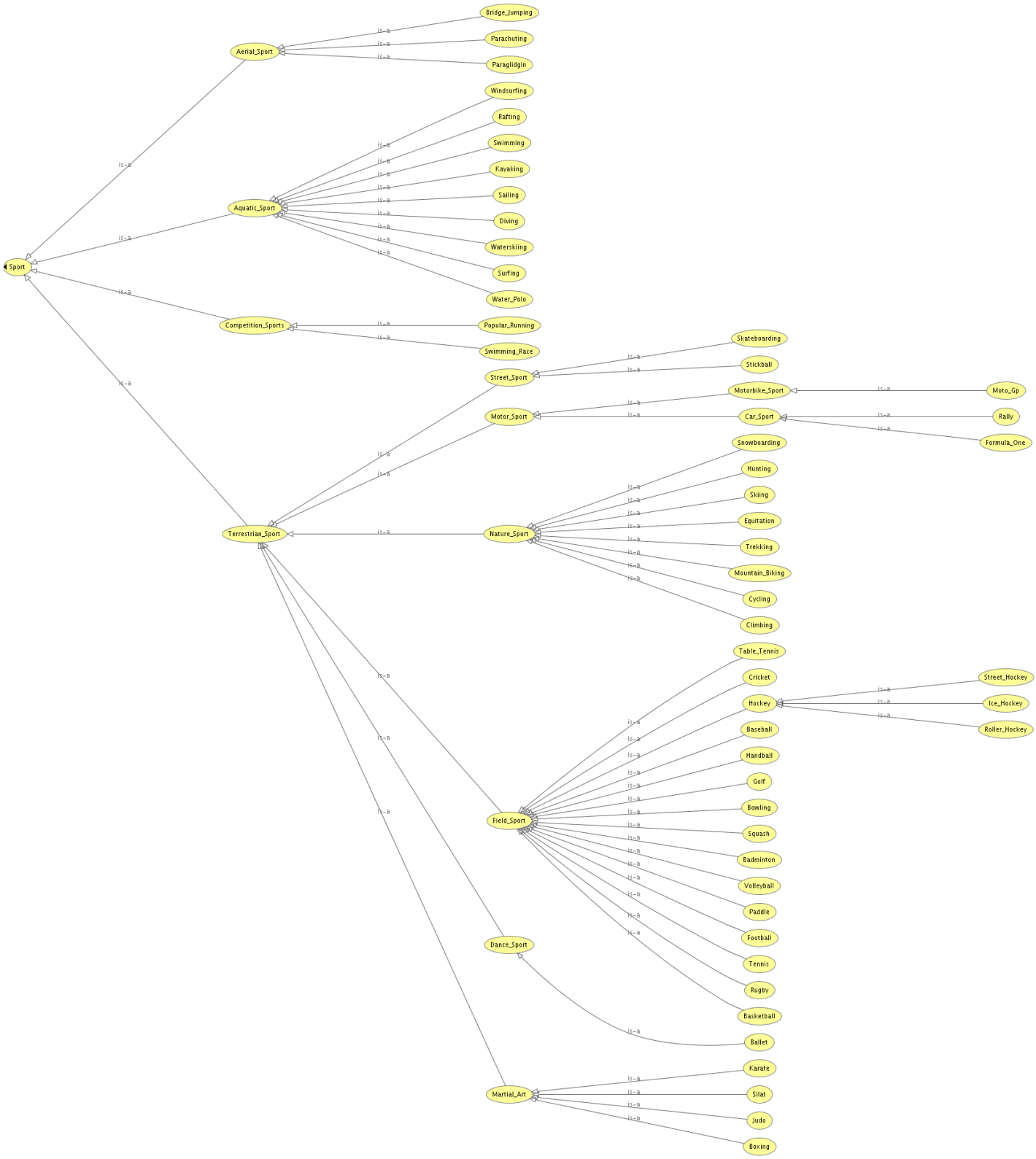
4 Annex I – TourismOWL

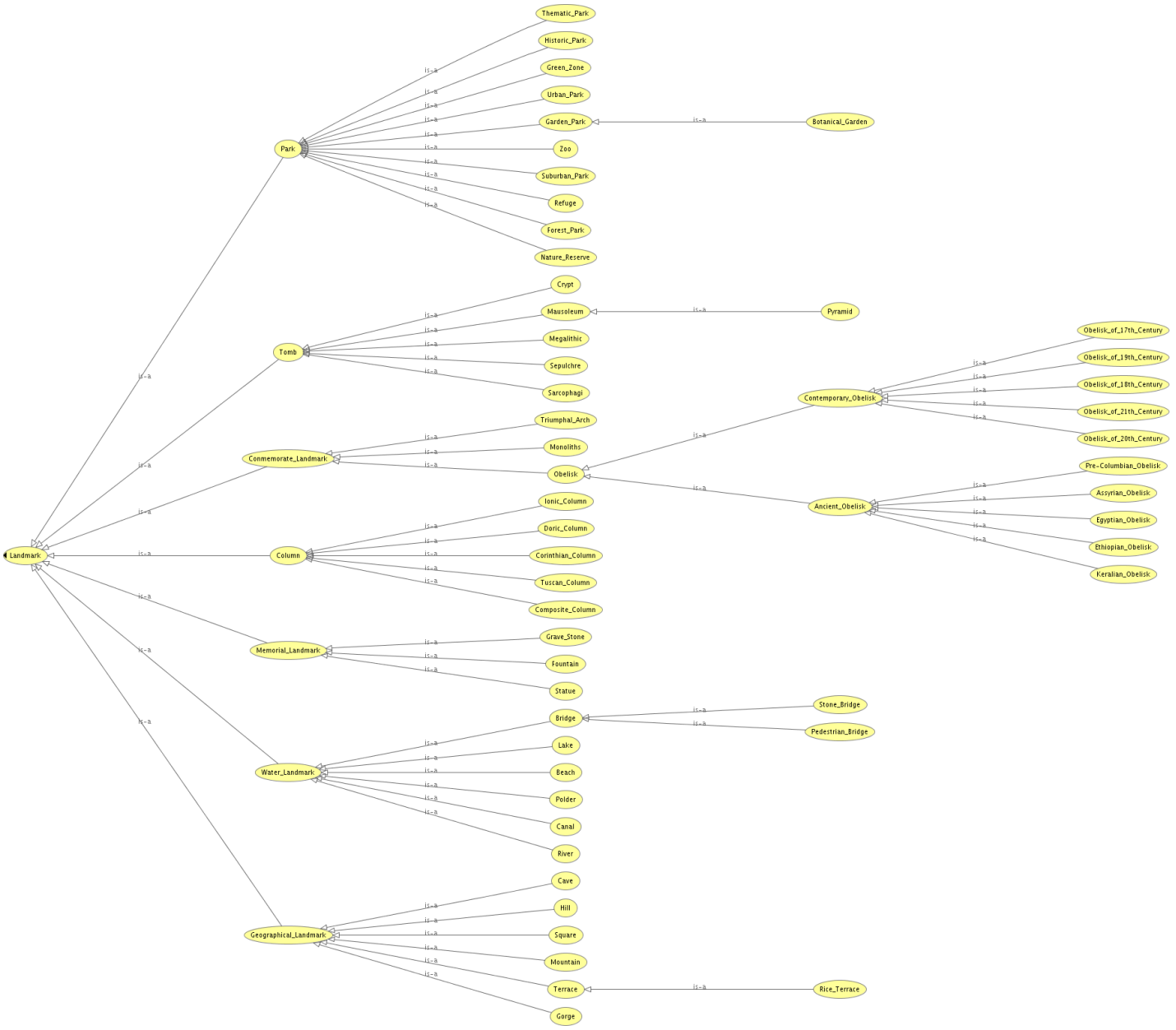
Following it is shown the taxonomy of the TourismOWL.owl ontology. Each sub-tree depicts the main classes of the ontology and its hierarchy.


Museum
Art_Museum
Maritime_Museum
Science_Museum
Wax_Museum
Specialized_Museum
Virtual_Museum
Egyptian_Museum
Sex_Museum
Natural_History_Museum
Open_Air_Museum
Archeology_Museum
Mobile_Museum
Contemporary_Art_Museum
Folk_Art_Museum
Modern_Art_Museum
Art_Gallery
Industrial_Museum
Railway_Museum
Aviation_Museum
Astronomy_Museum
Computer_Museum
Technology_Museum
Physics_Museum
Children_Museum
Toy_Museum
Feminist_Museum
Biographical_Museum
Fishing_Museum
Music_Museum
Military_Museum
Woman_Museum
Erotic_Museum
Ecomuseum
Car_Museum
is-a


Festival
Christian_Festival
Music_Festival
Art_Festival
Song_Festival
Carnival
is-a

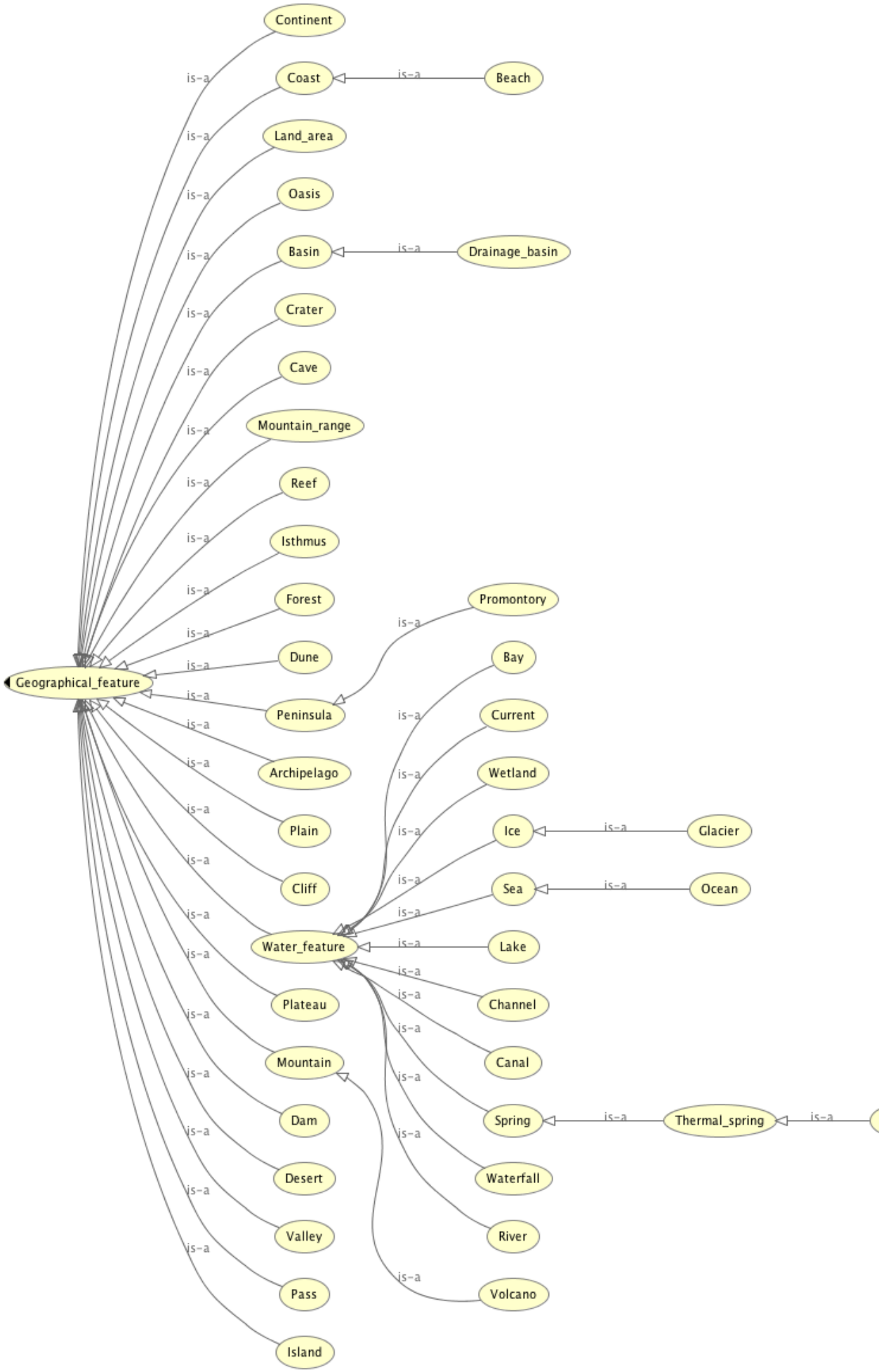

Building
is-a
Skyscraper
Industrial_Building
Comercial_Building
Sport_Building
Residential_Building
Cultural_Building
Religious_Building
Militar_Building
Textil_Mill
Cotton_Mill
Fair
Fashion_Fair
Artisan_Fair
Trade_Fair
Shopping
Shopping_Street
Shopping_Area
Shopping_Centre
Mall
Tiangui
Comercial_District
Comercial_Complex
Kiosk
Store
Convenience_Store
Boockstore
Clothes_Shop
Outlet
Fashion_Shop
Shop
Shophouse
Souvenir_Shop
Antiquarian_Shop
Market
Food_Market
Flea_Market
Hypermarket
Supermarket
Casino
Casino_Resort
Aquatic_Centre
Pool
Sport_Complex
Bowling_Alley
Squash_Centre
Golf_Course
Formula_One_Circuit
Velodrome
Stadium
Football_Stadium
Rugby_Stadium
Criket_Stadium
Hockey_Stadium
Baseball_Stadium
Camping
Residential_District
House
Residential_Tower
Townhouses
Hotel
Garden_Apartments
Luxury_Hotel
Ancient_Greek_Theatre
Theatre
University
Public_University
Technological_University
Private_University
Forum
School
Public_School
Private_School
Business_School
Art_School
Music_School
Amphitheatre
Library
Opera
Contemporany_Amphitheatre
Natural_Amphitheatre
Roman_Amphitheatre
Christian_Building
Chapel
Convent
Abbey
Church
Cathedral
Jewish_Building
Synagogue
Islamic_Building
Mosque
Miscellaneous_Building
Basilica
Sanctuary
Temple
Parish
Monastery
Tower
Prison
Headquarter
Palace
Royal_Residence
Fort


Sport
Aerial_Sport
Bridge_Jumping
Parachuting
Paraglidgin
Aquatic_Sport
Windsurfing
Rafting
Swimming
Kayaking
Sailing
Diving
Waterskiing
Surfing
Water_Polo
Competition_Sports
Popular_Running
Swimming_Race
Terrestrian_Sport
Street_Sport
Skateboarding
Stickball
Motor_Sport
Motorbike_Sport
Moto_Gp
Car_Sport
Rally
Formula_One
Nature_Sport
Snowboarding
Hunting
Skiing
Equitation
Trekking
Mountain_Biking
Cycling
Climbing
Field_Sport
Table_Tennis
Cricket
Hockey
Street_Hockey
Ice_Hockey
Roller_Hockey
Baseball
Handball
Golf
Bowling
Squash
Badminton
Volleyball
Paddle
Football
Tennis
Rugby
Basketball
Dance_Sport
Ballet
Martial_Art
Karate
Silat
Judo
Boxing
is-a

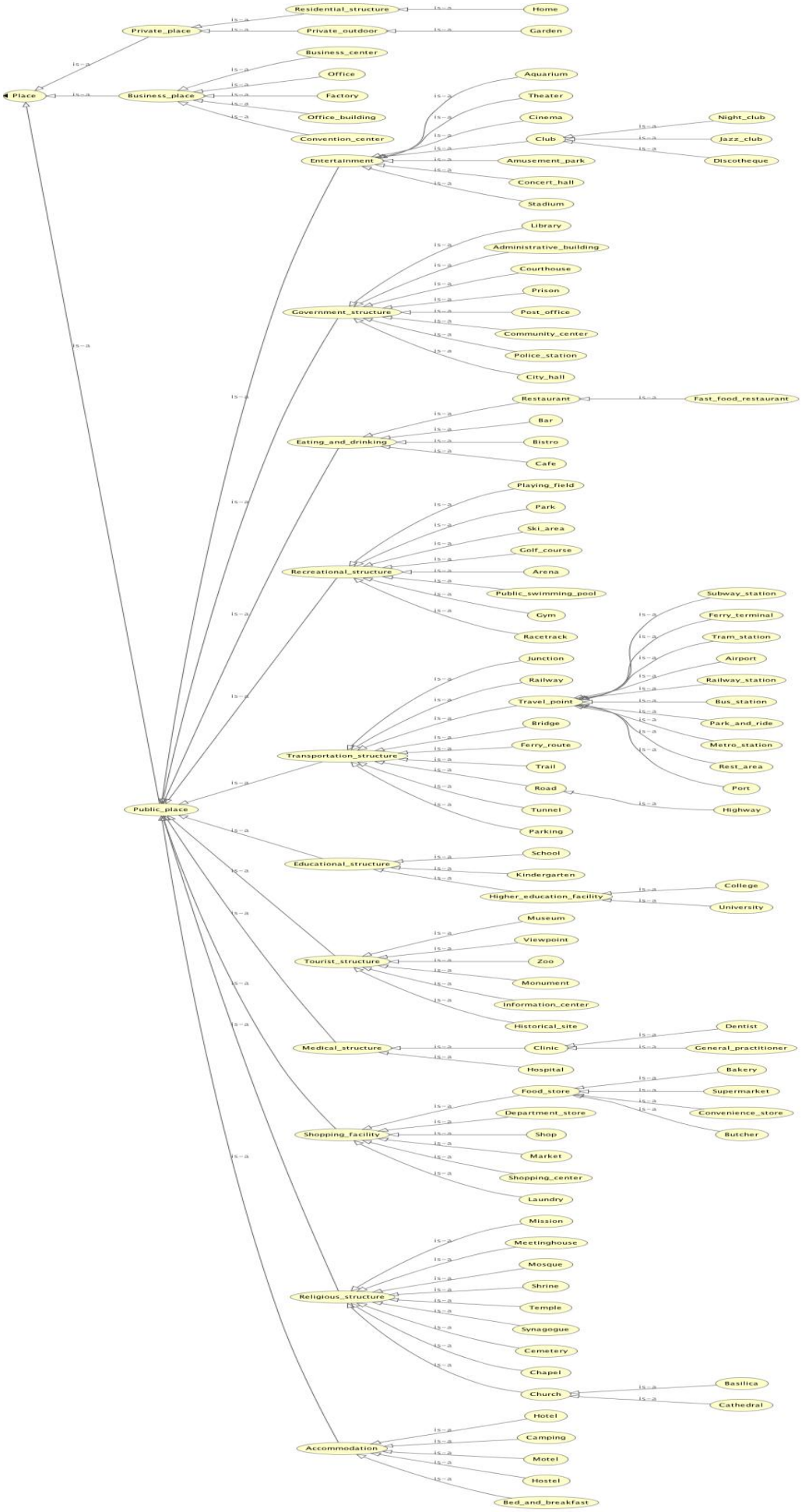

Landmark
is-a
Park
Thematic_Park
Historic_Park
Green_Zone
Urban_Park
Garden_Park
Botanical_Garden
Zoo
Suburban_Park
Refuge
Forest_Park
Nature_Reserve
Tomb
Crypt
Mausoleum
Pyramid
Megalithic
Sepulchre
Sarcophagi
Conmemorate_Landmark
Triumphal_Arch
Monoliths
Obelisk
Contemporary_Obelisk
Obelisk_of_17th_Century
Obelisk_of_19th_Century
Obelisk_of_18th_Century
Obelisk_of_21th_Century
Obelisk_of_20th_Century
Ancient_Obelisk
Pre-Columbian_Obelisk
Assyrian_Obelisk
Egyptian_Obelisk
Ethiopian_Obelisk
Keralian_Obelisk
Column
Ionic_Column
Doric_Column
Corinthian_Column
Tuscan_Column
Composite_Column
Memorial_Landmark
Grave_Stone
Fountain
Statue
Bridge
Stone_Bridge
Pedestrian_Bridge
Water_Landmark
Lake
Beach
Polder
Canal
River
Geographical_Landmark
Cave
Hill
Square
Mountain
Terrace
Rice_Terrace
Gorge

5 Annex II – Space.owl

Following it is shown the taxonomy of the Space.owl ontology. Each sub-tree depicts the main classes of the ontology and its hierarchy.


Continent
Coast
Beach
Land_area
Oasis
Basin
Drainage_basin
Crater
Cave
Mountain_range
Reef
Isthmus
Forest
Dune
Geographical_feature
Peninsula
Promontory
Archipelago
Plain
Cliff
Plateau
Mountain
Dam
Desert
Valley
Pass
Island
Bay
Current
Wetland
Ice
Glacier
Sea
Ocean
Water_feature
Lake
Channel
Canal
Spring
Thermal_spring
Waterfall
River
Volcano
is-a

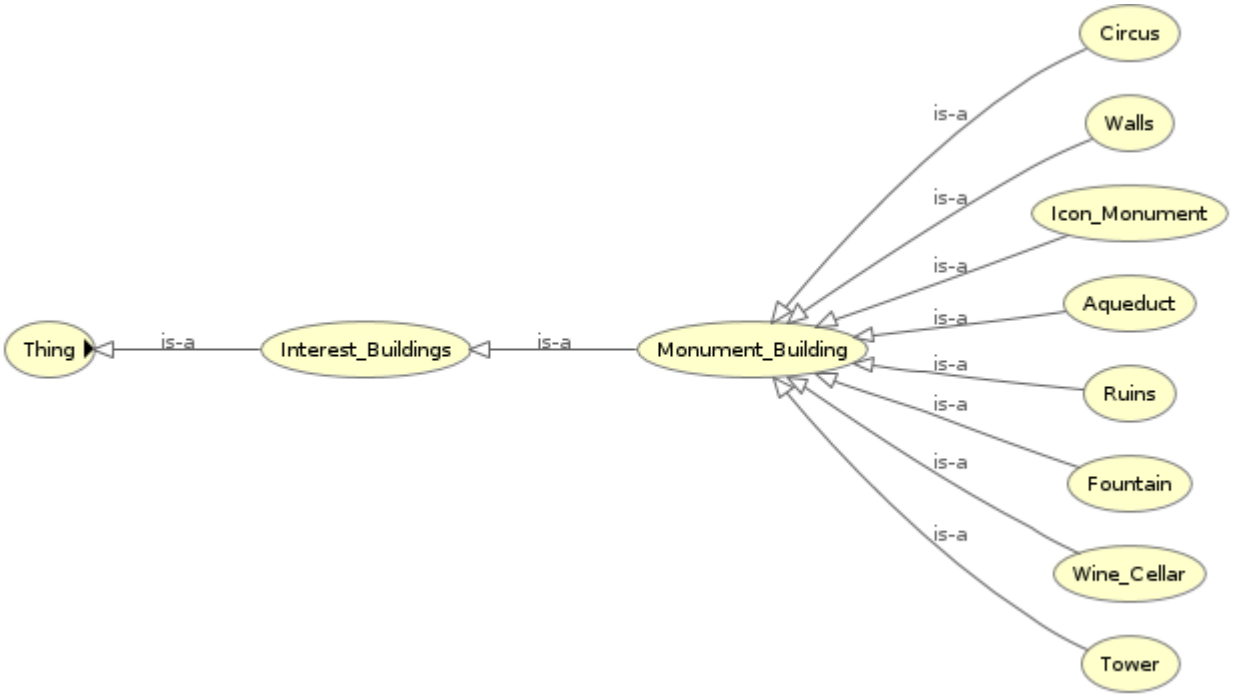

Geopolitical_entity
is-a
Subdivision
Prefecture
Zone
Department
Street
Locality
District
Province
Area
Canton
Parish
Autonomous_Community
County
Country
Republic
State
Monarchy
Emirate
Settlement
Hamlet
Capital
City
Village
Megalopolis
Town
Metropolis

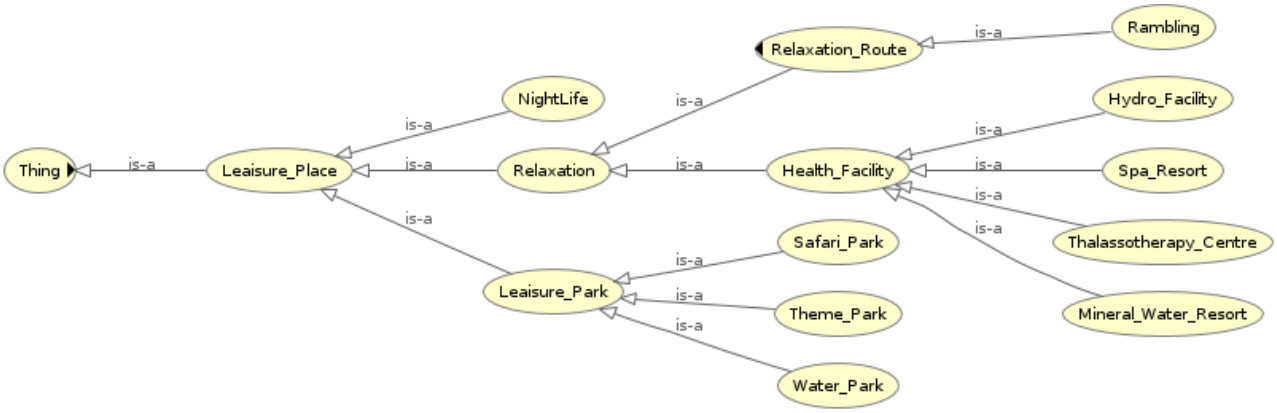

Place
Private_place
Residential_structure
Home
Private_outdoor
Garden
Business_place
Business_center
Office
Factory
Office_building
Convention_center
Public_place
Entertainment
Aquarium
Theater
Cinema
Club
Night_club
Jazz_club
Discotheque
Amusement_park
Concert_hall
Stadium
Government_structure
Library
Administrative_building
Courthouse
Prison
Post_office
Community_center
Police_station
City_hall
Eating_and_drinking
Restaurant
Fast_food_restaurant
Bar
Bistro
Cafe
Recreational_structure
Playing_field
Park
Ski_area
Golf_course
Arena
Public_swimming_pool
Gym
Racetrack
Transportation_structure
Junction
Railway
Travel_point
Subway_station
Ferry_terminal
Tram_station
Airport
Railway_station
Bus_station
Park_and_ride
Metro_station
Rest_area
Port
Bridge
Ferry_route
Trail
Road
Highway
Tunnel
Parking
Educational_structure
School
Kindergarten
Higher_education_facility
College
University
Tourist_structure
Museum
Viewpoint
Zoo
Monument
Information_center
Historical_site
Medical_structure
Clinic
Dentist
General_practitioner
Hospital
Shopping_facility
Food_store
Bakery
Supermarket
Convenience_store
Butcher
Department_store
Shop
Market
Shopping_center
Laundry
Religious_structure
Mission
Meetinghouse
Mosque
Shrine
Temple
Synagogue
Cemetery
Chapel
Church
Basilica
Cathedral
Accommodation
Hotel
Camping
Motel
Hostel
Bed_and_breakfast
is-a

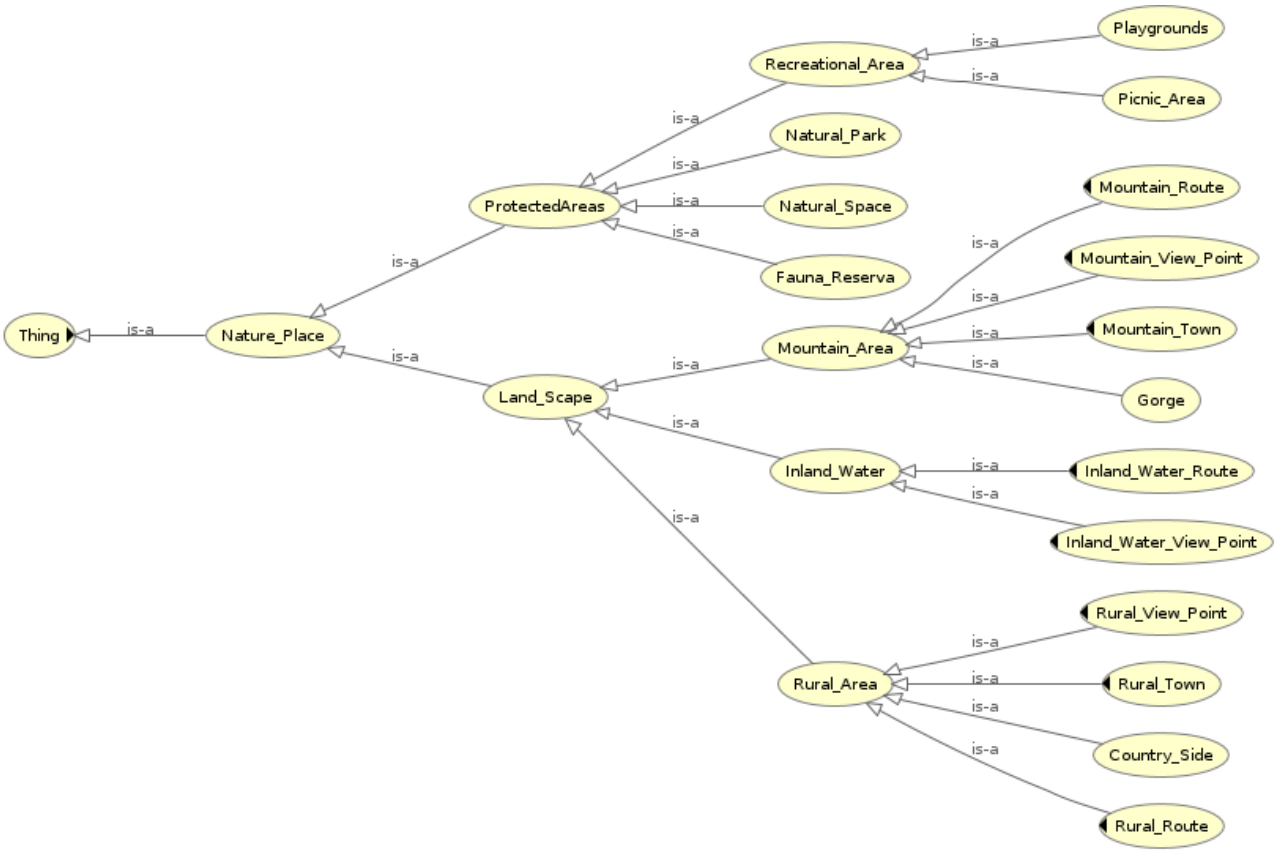
6 Annex III – PCTTO.owl

Following it is shown the taxonomy of the PCTTO.owl ontology. Each sub-tree depicts the main classes of the ontology and its hierarchy.


Rappelling
Gorge_Walking
Bungee_Jumping
Spelunking
Zip_Line
Raids
Orienting_Course
AdventureSports
Thing
is-a
Meaning_Landscape
Urban_Landscape
Culture_Landscape
Book_Faire
Wine_Event
Gastronomy_Event
Food_Event
Traditional_Celebration
Dance_Festival
Sport_Event
Arts_And_Crafts_Event
Popular_Celebration
Event
Hang_Gliding
Archery
Clay_Pigeon_Shooting
PaintBall
Shooting_Sports


Thing
is-a
Interest_Buildings
is-a
Monument_Building
is-a
Circus
is-a
Walls
is-a
Icon_Monument
is-a
Aqueduct
is-a
Ruins
is-a
Fountain
is-a
Wine_Cellar
is-a
Tower

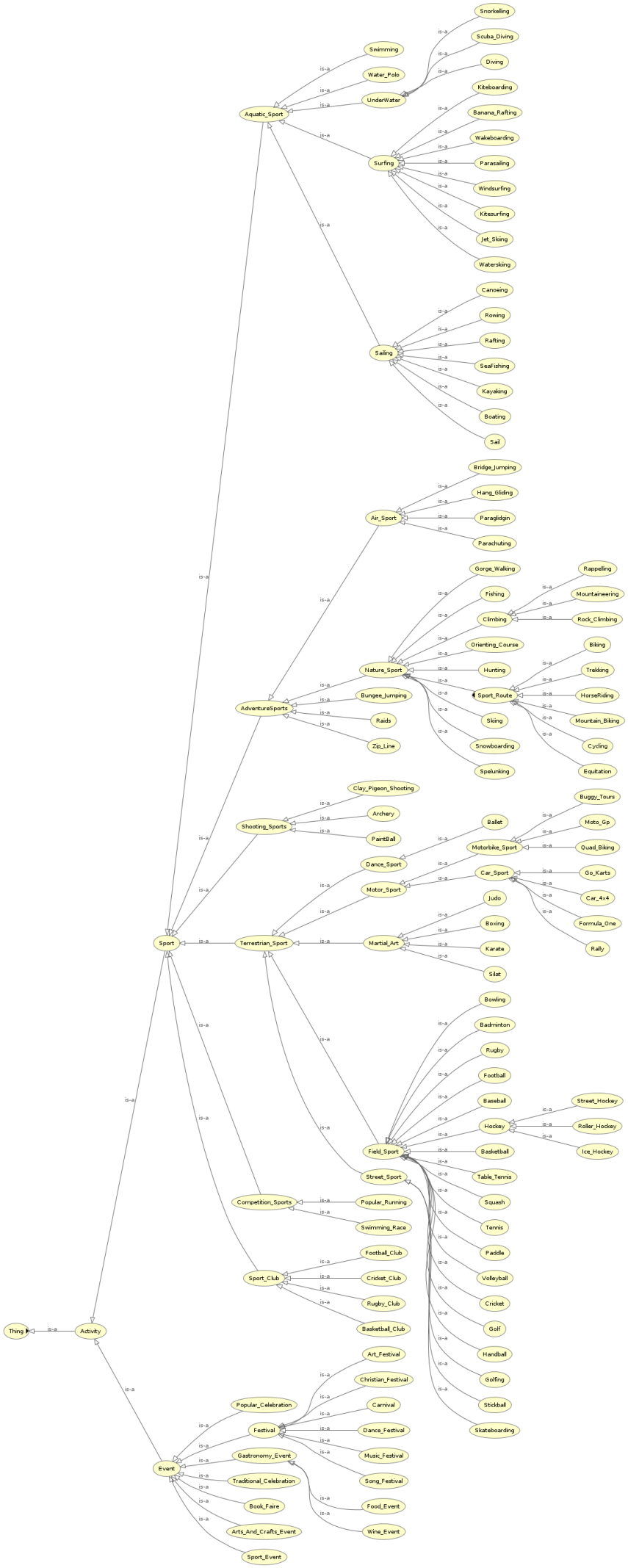

Thing
is-a
Leaisure_Place
is-a
NightLife
is-a
Relaxation
is-a
Relaxation_Route
is-a
Rambling
is-a
Health_Facility
is-a
Hydro_Facility
is-a
Spa_Resort
is-a
Thalassotherapy_Centre
is-a
Mineral_Water_Resort
is-a
Leaisure_Park
is-a
Safari_Park
is-a
Theme_Park
is-a
Water_Park

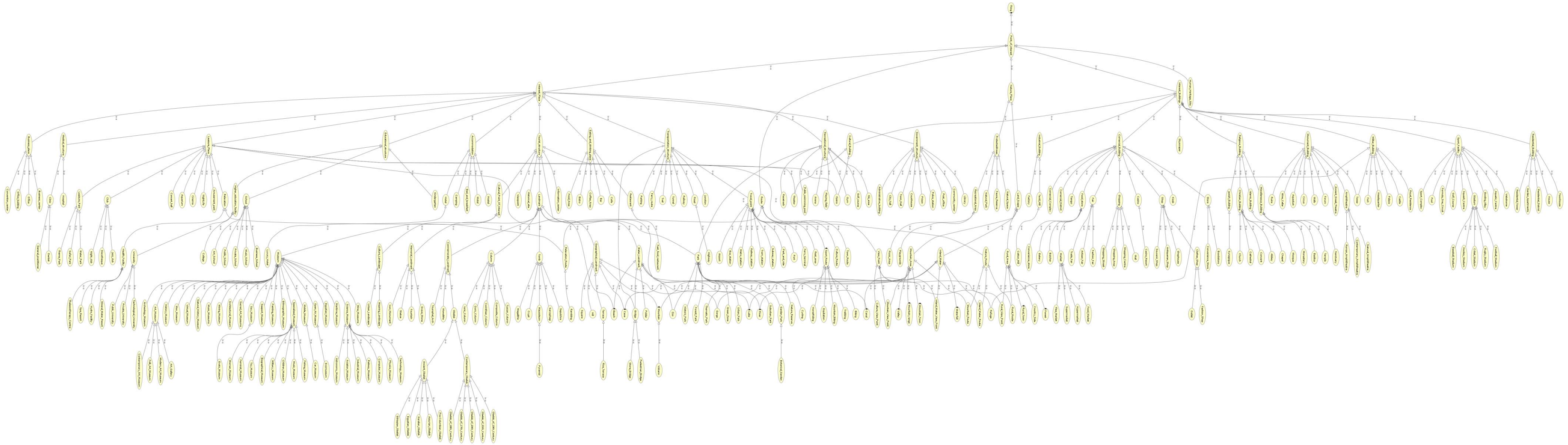

Thing
is-a
Nature_Place
is-a
ProtectedAreas
is-a
Recreational_Area
is-a
Playgrounds
is-a
Picnic_Area
is-a
Natural_Park
is-a
Natural_Space
is-a
Fauna_Reserva
is-a
Land_Scape
is-a
Mountain_Area
is-a
Mountain_Route
is-a
Mountain_View_Point
is-a
Mountain_Town
is-a
Gorge
is-a
Inland_Water
is-a
Inland_Water_Route
is-a
Inland_Water_View_Point
is-a
Rural_Area
is-a
Rural_View_Point
is-a
Rural_Town
is-a
Country_Side
is-a
Rural_Route

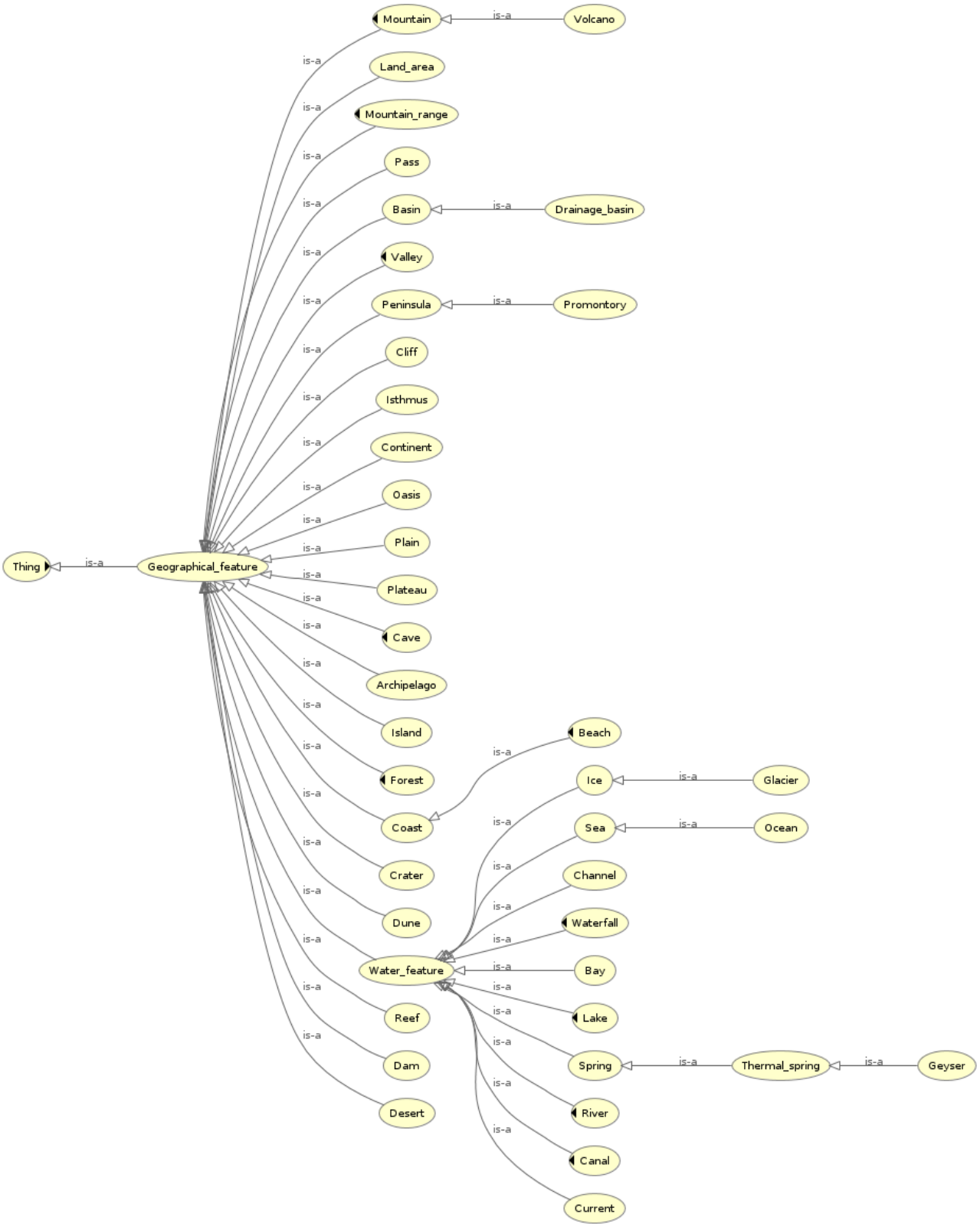

Thing
is-a
Route
Nature_Route
Rural_Route
Mountain_Route
Inland_Water_Route
Sport_Route
Biking
Trekking
HorseRiding
Culture_Route
Town_Route
Relaxation_Route
Rambling
Sailing
Boating
SeaFishing
Sail
Canoeing
Rowing
Towns
Traditional_Town
Rural_Town
Town_Route
Mountain_Town
UnderWater
Snorkelling
Scuba_Diving
View_Point
Mountain_View_Point
Culture_View_Point
Inland_Water_View_Point
Rural_View_Point

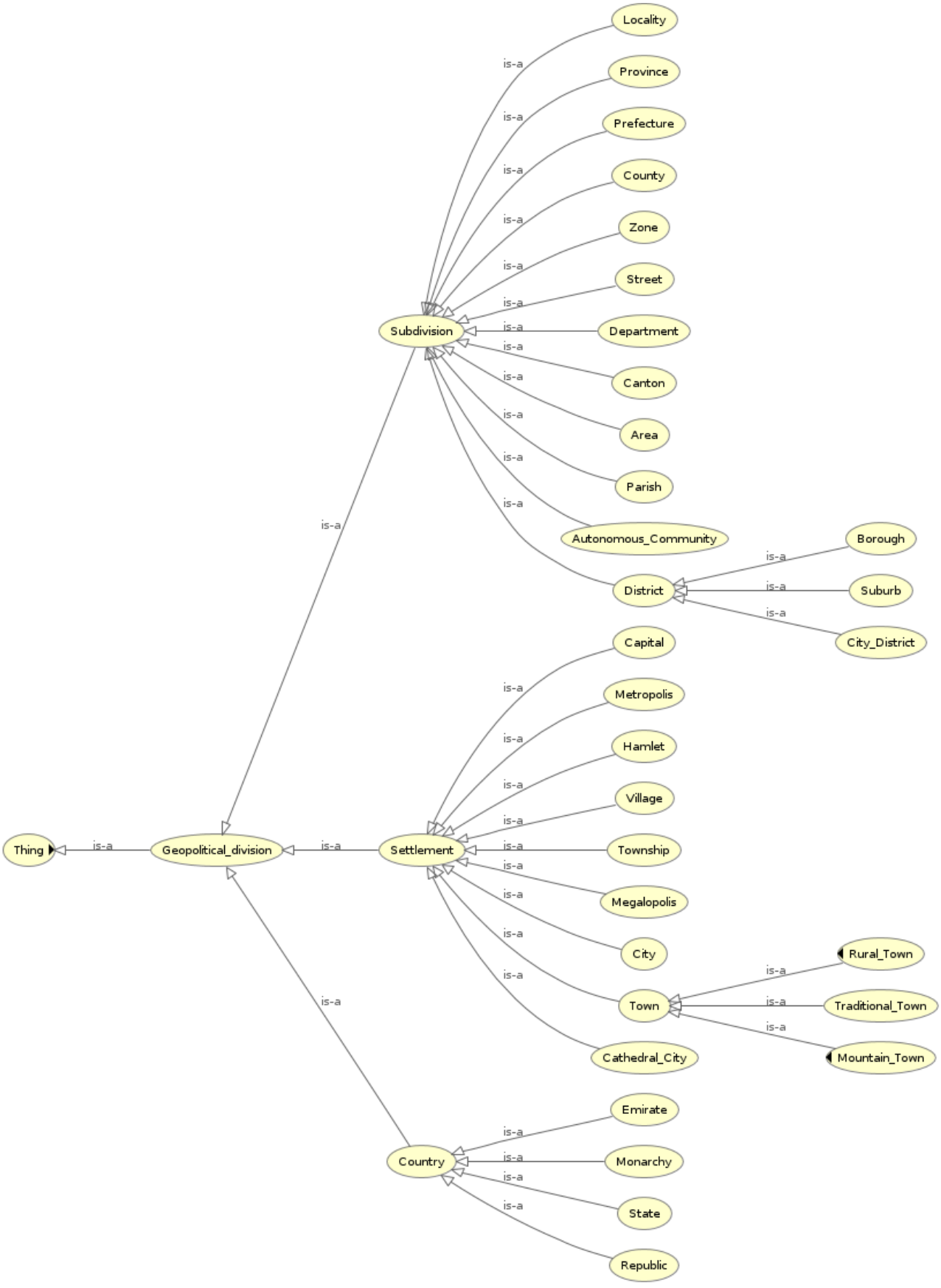
7 Annex IV – damaskontology.owl

Following it is shown the taxonomy of the damaskontology.owl ontology. Each sub-tree depicts the main classes of the ontology and its hierarchy.


Thing
is-a
Activity
Event
Festival
Popular_Celebration
Gastronomy_Event
Traditional_Celebration
Book_Faire
Arts_And_Crafts_Event
Sport_Event
Art_Festival
Christian_Festival
Carnival
Dance_Festival
Music_Festival
Song_Festival
Food_Event
Wine_Event
Sport
Aquatic_Sport
Swimming
Water_Polo
UnderWater
Snorkelling
Scuba_Diving
Diving
Surfing
Kiteboarding
Banana_Rafting
Wakeboarding
Parasailing
Windsurfing
Kitesurfing
Jet_Skiing
Waterskiing
Sailing
Canoeing
Rowing
Rafting
SeaFishing
Kayaking
Boating
Sail
Air_Sport
Bridge_Jumping
Hang_Gliding
Paraglidgin
Parachuting
AdventureSports
Bungee_Jumping
Raids
Zip_Line
Nature_Sport
Gorge_Walking
Fishing
Climbing
Rappelling
Mountaineering
Rock_Climbing
Orienting_Course
Hunting
Sport_Route
Biking
Trekking
HorseRiding
Mountain_Biking
Cycling
Equitation
Skiing
Snowboarding
Spelunking
Shooting_Sports
Clay_Pigeon_Shooting
Archery
PaintBall
Terrestrian_Sport
Dance_Sport
Ballet
Motor_Sport
Motorbike_Sport
Buggy_Tours
Moto_Gp
Quad_Biking
Car_Sport
Go_Karts
Car_4x4
Formula_One
Rally
Martial_Art
Judo
Boxing
Karate
Silat
Competition_Sports
Popular_Running
Swimming_Race
Sport_Club
Football_Club
Cricket_Club
Rugby_Club
Basketball_Club
Field_Sport
Bowling
Badminton
Rugby
Football
Baseball
Hockey
Street_Hockey
Roller_Hockey
Ice_Hockey
Basketball
Table_Tennis
Squash
Tennis
Paddle
Volleyball
Cricket
Golf
Handball
Golfing
Stickball
Skateboarding
Street_Sport




Thing
is-a
Geographical_feature
is-a
Mountain
Volcano
Land_area
Mountain_range
Pass
Basin
Drainage_basin
Valley
Peninsula
Promontory
Cliff
Isthmus
Continent
Oasis
Plain
Plateau
Cave
Archipelago
Island
Forest
Coast
Beach
Crater
Dune
Water_feature
Ice
Glacier
Sea
Ocean
Channel
Waterfall
Bay
Lake
Spring
Thermal_spring
Geyser
River
Canal
Current
Reef
Dam
Desert


Thing
is-a
Geopolitical_division
is-a
Settlement
is-a
Subdivision
is-a
Locality
Province
Prefecture
County
Zone
Street
Department
Canton
Area
Parish
Autonomous_Community
District
Borough
Suburb
City_District
Capital
Metropolis
Hamlet
Village
Township
Megalopolis
City
Town
Rural_Town
Traditional_Town
Mountain_Town
Cathedral_City
is-a
Country
Emirate
Monarchy
State
Republic