

The GeoLOD catalog and recommender for spatial linked data – A demonstration

Vasilis Kopsachilis and Michail Vaitis

Department of Geography, University of the Aegean, Greece



UNIVERSITY OF THE AEGEAN

30 May 2022 in Hersonissos, Greece

GeoLOD

<http://geolod.net>

- A web catalog of semantic web spatial datasets and classes (datasets and classes that contain georeferenced instances)
- A recommender for relevant datasets and classes for link discovery processes (datasets and classes that contain related instances, such as sameAs)

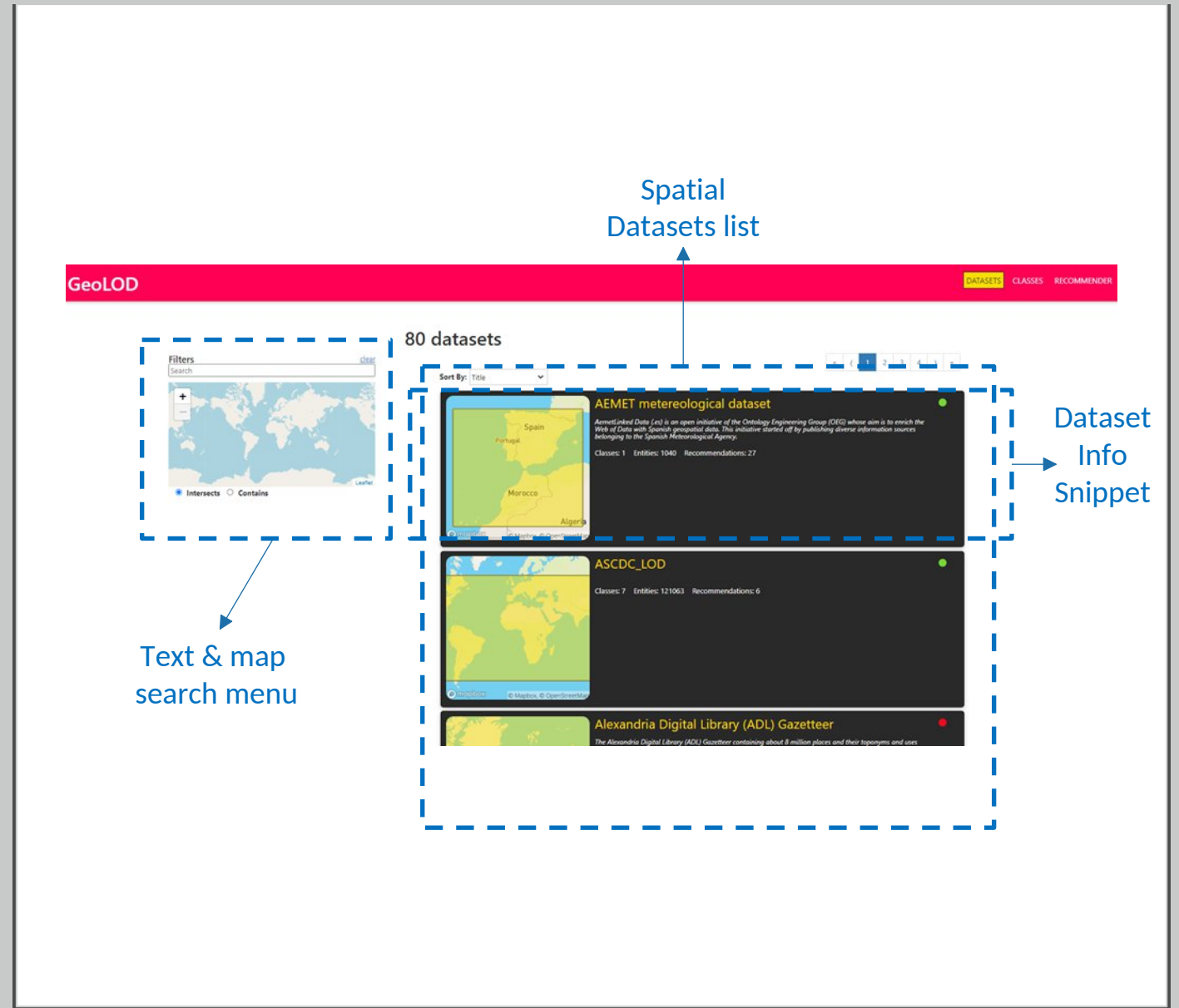
The screenshot displays the GeoLOD web application interface. At the top, a pink header bar contains the "GeoLOD" logo on the left and navigation links "DATASETS", "CLASSES", and "RECOMMENDER" on the right. Below the header, the main content area is titled "80 datasets". On the left side of this area, there is a "Filters" section with a search bar and a world map. The map has a legend with "Intersects" (selected) and "Contains" options. To the right of the filters, a "Sort By:" dropdown menu is set to "Title". Below the sorting menu, a list of datasets is shown. The first dataset is "AEMET meteorological dataset", which includes a map of Spain and Morocco, a description of the dataset, and statistics: "Classes: 1", "Entities: 1040", and "Recommendations: 27". The second dataset is "ASCDC_LOD", with a world map, description, and statistics: "Classes: 7", "Entities: 121063", and "Recommendations: 6". The third dataset is "Alexandria Digital Library (ADL) Gazetteer", with a map of Egypt, description, and statistics: "Classes: 1", "Entities: 8000000", and "Recommendations: 1". Each dataset entry has a green dot in the top right corner, indicating it is a recommended dataset.

GeoLOD highlights

- Focuses on the geographic semantic web exploration
- Exploits dataset spatial characteristics to aid exploration
- Creates and preserves spatial metadata for datasets, exposed in GeoVoID, a dataset metadata vocabulary that extends VoID
- Offers text and map-based dataset & classes search (e.g., find classes that contain instances in Crete)
- Maintains pre-computed recommendation lists for classes that may contain related instances (e.g., sameAs) to aid link discovery
- Offers on-the-fly recommendations for uncatalogued datasets and GIS spatial datasets (Shapefile, GeoJSON)
- Exports of Silk and LIMES configuration files for recommended pair of classes for direct use in link discovery processes
- Offers a REST API for software-based consumption of its content

GeoLOD Catalog

- Parses the LOD cloud and the Datahub & locates SPARQL Endpoints (datasets) that contain georeferenced instances
- Extracts datasets and classes basic metadata and generates additional spatial metadata (Bounding Box, number of georeferenced instances, spatial vocabularies)
- Provide access to lists of spatial datasets and classes through a user interface that allows text and map-based search



Spatial Dataset Collector

1. Finds available SPARQL Endpoints by parsing:
 - The list of LOD datasets exposed in JSON (<https://lod-cloud.net/lod-data.json>)
 - The list of DatatHub datasets exposed in CKAN API (<https://old.datahub.io/>)
2. Identifies SPARQL Endpoints that contain georeferenced point instances with one of the W3C Basic Geo, GeoSPARQL, GeoVocab, Geonames and GeoRSS spatial vocabularies by sending SPARQL ASK queries.

W3C Basic Geo ASK query

```
ASK {  
  ?s <http://www.w3.org/2003/01/geo/wgs84_pos#long> ?x.  
  ?s <http://www.w3.org/2003/01/geo/wgs84_pos#lat> ?y.  
}
```

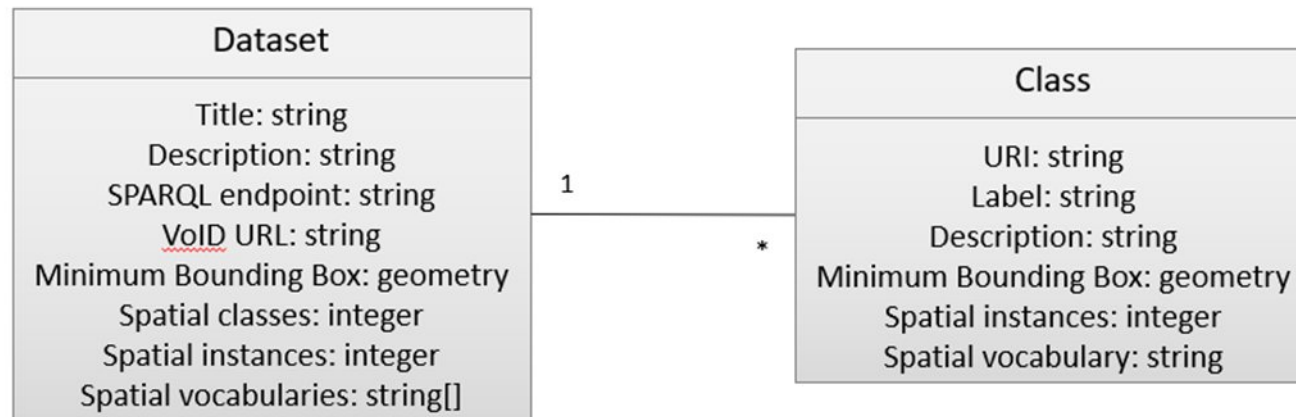
3. Retrieves the spatial classes found in each SPARQL Endpoint by sending SPARQL SELECT queries

List of classes containing instances that use W3C Basic Geo

```
SELECT DISTINCT ?class {  
  ?s <http://www.w3.org/2003/01/geo/wgs84_pos#long> ?x.  
  ?s <http://www.w3.org/2003/01/geo/wgs84_pos#lat> ?y.  
  ?s <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> ?class.  
}
```

Dataset & Class Metadata

- Basic metadata extracted by LOD and Datahub:
 - For datasets ☾ title, description, SPARQL endpoint URL and VoID URL
 - For classes ☾ URI, label, description
- Additional “spatial” metadata extracted by sending SPARQL SELECT queries to endpoints:
 - For datasets ☾ Bounding Box, # of spatial classes & instances, spatial vocabularies
 - For classes ☾ Bounding Box, # of instances, spatial vocabularies
- Schema of GeoLOD metadata database:



Dataset View

GeoLOD

DATASETS
CLASSES
RECOMMENDER

AEMET metereological dataset



AemetLinked Data (.es) is an open initiative of the Ontology Engineering Group (OEG) whose aim is to enrich the Web of Data with Spanish geospatial data. This initiative started off by publishing diverse information sources belonging to the Spanish Meteorological Agency.

SPARQL Endpoint: <http://aemet.linkeddata.es/sparql>

Spatial Ontologies: W3C Basic Geo (http://www.w3.org/2003/01/geo/wgs84_pos)

Entities: 780

Classes: 1

Recommendations: 27

Dataset Metadata

Download VOID
Export GeoVOID
Export Recommendations

Various actions

Classes

10 ▾ Showing 1 to 1 of 1 classes

1

Label ▾▲	URI ▾▲	Ontology ▾▲	Entities ▾▲	Recommendations ▾▲
Enter Label...				
Point	http://www.w3.org/2003/01/geo/wgs84_pos#Point	W3C Basic Geo	260	27

Dataset Classes

GeoVoID

A vocabulary that extends VoID for describing spatial datasets and classes.

It defines 3 datatype properties:

- **geovoid:entities**  The total number of spatial entities that are described in a void:Dataset or void:Class
- **geovoid:classes**  The total number of distinct spatial classes in a void:Dataset
- **geovoid:vocabulary**  A spatial vocabulary that is used in the dataset

Schema definition is online at:

<http://83.212.121.124/schemas/geovoid#>

GeoVoID description of the AEMET dataset

```
@prefix geovoid: <http://83.212.121.124/schemas/geovoid#>.
@prefix void: <http://rdfs.org/ns/void#>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>.
@prefix dcterms: <http://purl.org/dc/terms/>.
@prefix locn: <https://www.w3.org/ns/locn#>.

<AEMET> a void:Dataset;
    dcterms:title "AEMET metereological dataset";
    dcterms:description "AemetLinked Data (.es) is an open initiative of
geospatial data. This initiative started off by publishing diverse inform
void:sparqlEndpoint <http://aemet.linkeddata.es/sparql>;
    dcterms:spatial [
    a dcterms:Location;
    locn:geometry "POLYGON((-17.9161111111 27.7355555556,-17.9161111111 43
];
    geovoid:vocabulary <http://www.w3.org/2003/01/geo/wgs84_pos>;
    geovoid:entities "1300"^^xsd:integer;
    geovoid:classes "1"^^xsd:integer;
    void:classPartition [
    void:class <http://www.w3.org/2003/01/geo/wgs84_pos#Point>;
    dcterms:spatial [
    a dcterms:Location;
    locn:geometry "POLYGON((-17.9161111111 27.7355555556,4.2155555556 27.
];
    geovoid:vocabulary <http://www.w3.org/2003/01/geo/wgs84_pos>;
    geovoid:entities "260"^^xsd:integer
].
```

Class View

GeoLOD

DATASETS

CLASSES



mapbox © Mapbox, © OpenStreetMap

Point

http://www.w3.org/2003/01/geo/wgs84_pos#Point

Dataset: AEMET meteorological dataset

Spatial Ontology: W3C Basic Geo (http://www.w3.org/2003/01/geo/wgs84_pos)

Entities: 260

Recommendations: 33

Download RDF

Download JSON

Download GeoJSON

View On Map

Class Metadata

Various actions

Recommendations

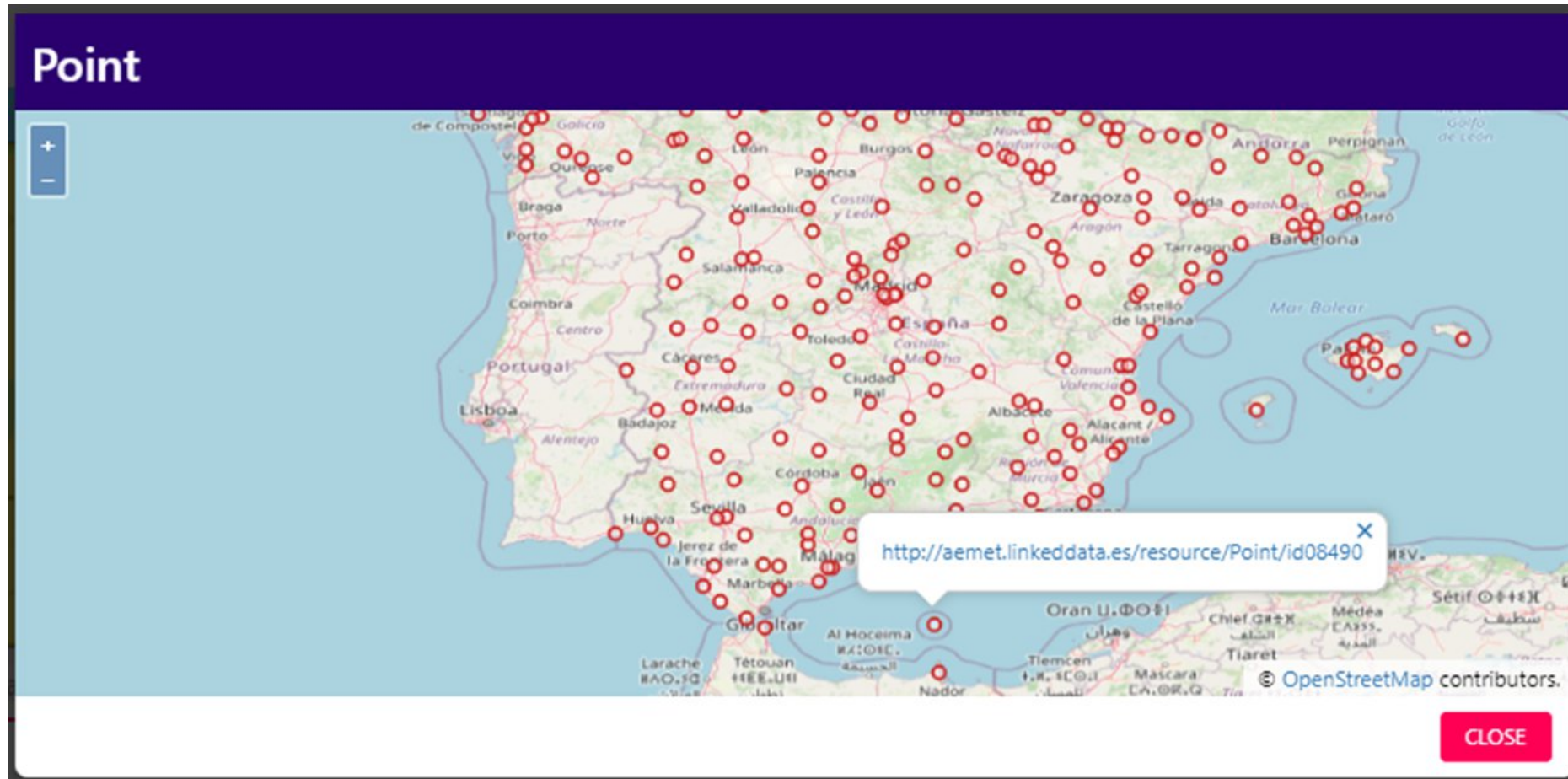
10 Showing 1 to 10 of 33 classes

1 2 3 4 >

Label	URI	Dataset	Est. Related Instances	Rank	Link Discovery
Windsock	http://linkedgeodata.org/ontology/Windsock	LinkedGeoData	5	1	S L
Windsock	http://linkedgeodata.org/ontology/Windsock	LinkLion - A Link Repository for the Web of Data	4	2	S L

Class Recommendations

On-the-fly class instances preview on map



GeoLOD Recommender

- Recommends to a given class other classes that may contain related instances, such as sameAs.
- Maintains pre-computed recommendation lists for all classes in the catalog.
- Supports on-the-fly recommendations for uncatalogued SPARQL endpoints and GIS datasets (Shapefiles and GeoJSON).

GeoLOD DATASETS **CLASSES** RECOMMENDER

Recommendations

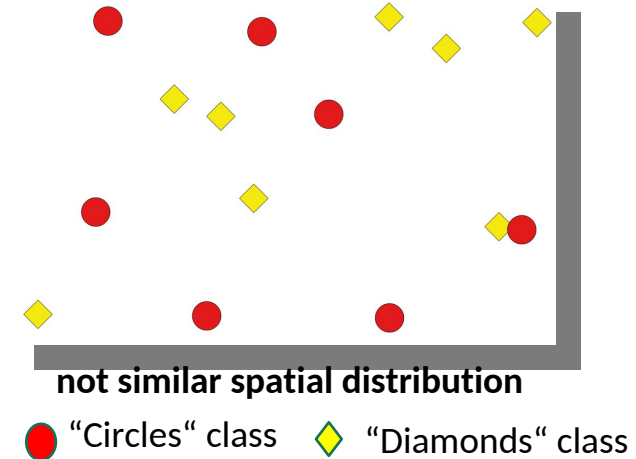
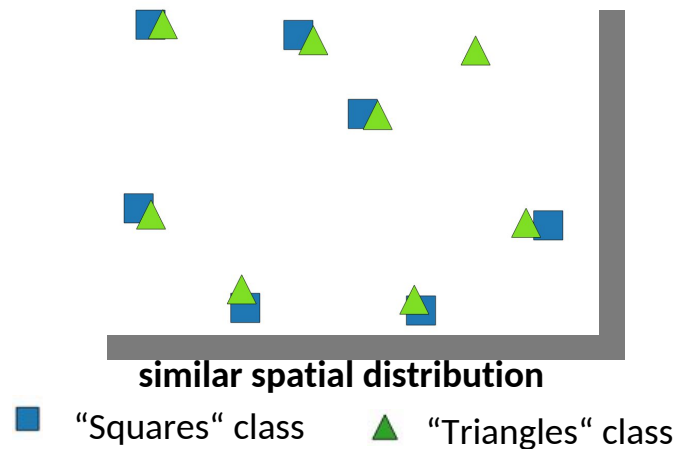
10 Showing 1 to 10 of 27 classes

Label	URI	Dataset	Common Entities	Rank	Link Discovery
DragLift	http://linkedgeodata.org/ontology/DragLift	LinkLion - A Link Repository for the Web of Data	3	1	S L
Windsock	http://linkedgeodata.org/ontology/Windsock	LinkLion - A Link Repository for the Web of Data	4	2	S L
Windsock	http://linkedgeodata.org/ontology/Windsock	LinkedGeoData	5	3	S L
AirportGate	http://linkedgeodata.org/ontology/AirportGate	LinkLion - A Link Repository for the Web of Data	4	4	S L
AirportGate	http://linkedgeodata.org/ontology/AirportGate	LinkedGeoData	4	5	S L
Beacon	http://linkedgeodata.org/ontology/Beacon	LinkLion - A Link Repository for the Web of Data	5	6	S L
Beacon	http://linkedgeodata.org/ontology/Beacon	LinkedGeoData	6	7	S L
ParkingSpace	http://linkedgeodata.org/ontology/ParkingSpace	LinkLion - A Link Repository for the Web of Data	3	8	S L
terms#Stop	http://vocab.gtfs.org/terms#Stop	Aragon Interoperable Information Structure E12A - aragon open data	12	9	S L
Lighthouse	http://linkedgeodata.org/ontology/Lighthouse	LinkLion - A Link Repository for the Web of Data	13	10	S L

Recommendation Algorithm

Based on the spatial distribution similarity of classes:

“pairs of classes whose instances present similar spatial distribution are more related than pairs of classes whose instances present dissimilar spatial distribution, in the sense that the former are more likely to contain semantically related instances”



In a nutshell, the algorithm builds and stores summaries for all identified spatial classes that capture their spatial extent and index the location of their spatial instances on a QuadTree, and then apply metrics that compute the similarity of the class summaries.

Recommendation List

Relevant classes to a “Meteorological Stations” class provided by the AEMET dataset

Recommendations					
10 ▾ Showing 1 to 10 of 33 classes		1 2 3 4 >			
Label ▾ Enter Label...	URI ▾	Dataset ▾ Enter Dataset...	Est. Related Instances ▾	Rank ▲	Link Discovery
Windsock	http://linkedgeodata.org/ontology/Windsock	LinkedGeoData	5	1	 
Windsock	http://linkedgeodata.org/ontology/Windsock	LinkLion - A Link Repository for the Web of Data	4	2	 
DragLift	http://linkedgeodata.org/ontology/DragLift	LinkLion - A Link Repository for the Web of Data	3	3	 
AerowayThing	http://linkedgeodata.org/ontology/AerowayThing	LinkLion - A Link Repository for the Web of Data	25	4	 
Lighthouse	http://linkedgeodata.org/ontology/Lighthouse	LinkLion - A Link Repository for the Web of Data	13	5	 
AerowayThing	http://linkedgeodata.org/ontology/AerowayThing	LinkedGeoData	24	6	 
Beacon	http://linkedgeodata.org/ontology/Beacon	LinkedGeoData	6	7	 
Lighthouse	http://linkedgeodata.org/ontology/Lighthouse	LinkedGeoData	13	8	 
UWWTP	http://rdfdata.eionet.europa.eu/uwwtd/ontology/UWWTP	European Pollutant Release and Transfer Register (E-PRTR)	51	9	 

Silk & LIMES linkage specification export

Default matching rules based on label Levenshtein and geometry Euclidean distances

Silk configuration file example

```
<Silk>
  <Prefixes>
    <Prefix id="rdf" namespace="http://www.w3.org/1999/02/22-rdf-syntax-ns#" />
    <Prefix id="owl" namespace="http://www.w3.org/2002/07/owl#" />
    <Prefix id="rdfs" namespace="http://www.w3.org/2000/01/rdf-schema#" />
    <Prefix id="xsd" namespace="http://www.w3.org/2001/XMLSchema#" />
  </Prefixes>
  <DataSources>
    <Dataset id="source" type="sparqlEndpoint">
      <Param name="endpointURI">http://aemet.linkeddata.es/sparql</Param>
    </Dataset>
    <Dataset id="target" type="sparqlEndpoint">
      <Param name="endpointURI">http://linkedgeodata.org/sparql</Param>
    </Dataset>
  </DataSources>
  <Interlinks>
    <Interlink id="l1">
      <SourceDataset dataSource="source" var="a" typeUri="http://www.w3.org/2003/01/geo/wgs84_pos#Point">
        <RestrictTo> ?a <http://www.w3.org/2003/01/geo/wgs84_pos#geometry> ?s . </RestrictTo>
      </SourceDataset>
      <TargetDataset dataSource="target" var="b" typeUri="http://linkedgeodata.org/ontology/Windsock">
        <RestrictTo> ?b <http://www.w3.org/2003/01/geo/wgs84_pos#geometry> ?t . </RestrictTo>
      </TargetDataset>
      <LinkageRule linkType="owl:sameAs">
        <Aggregate id="and" required="true" weight="1" type="min">
          <Compare id="levenshtein" required="true" weight="1" metric="levenshtein" threshold="0.2" indexing="true">
            <TransformInput id="trim1" function="trim">
              <TransformInput id="removeBlanks1" function="removeBlanks">
                <TransformInput id="removeParentheses1" function="removeParentheses">
                  <TransformInput id="removeSpecialChars1" function="removeSpecialChars">
                    <TransformInput id="alphaReduce1" function="alphaReduce">
                      <TransformInput id="lowerCase1" function="lowerCase">
                        <Input id="sourceLabel" path="?a/rdfs:label"/>
                      </TransformInput>
                    </TransformInput>
                  </TransformInput>
                </TransformInput>
              </TransformInput>
            </TransformInput>
          </Compare>
        </Aggregate>
      </LinkageRule>
    </Interlink>
  </Interlinks>
  <TransformInput id="trim2" function="trim">
    <TransformInput id="removeBlanks2" function="removeBlanks">
      <TransformInput id="removeParentheses2" function="removeParentheses">
        <TransformInput id="removeSpecialChars2" function="removeSpecialChars">
          <TransformInput id="alphaReduce2" function="alphaReduce">
            <TransformInput id="lowerCase2" function="lowerCase">
              <Input id="targetLabel" path="?b/rdfs:label"/>
            </TransformInput>
          </TransformInput>
        </TransformInput>
      </TransformInput>
    </TransformInput>
  </TransformInput>
```

LIMES configuration file example

```
<LIMES>
  <PREFIX>
    <NAMESPACE>http://www.w3.org/1999/02/22-rdf-syntax-ns#</NAMESPACE>
    <LABEL>rdf</LABEL>
  </PREFIX>
  <PREFIX>
    <NAMESPACE>http://www.w3.org/2002/07/owl#</NAMESPACE>
    <LABEL>owl</LABEL>
  </PREFIX>
  <PREFIX>
    <NAMESPACE>http://www.w3.org/2000/01/rdf-schema#</NAMESPACE>
    <LABEL>rdfs</LABEL>
  </PREFIX>
  <PREFIX>
    <NAMESPACE>http://www.w3.org/2003/01/geo/wgs84_pos#</NAMESPACE>
    <LABEL>geo</LABEL>
  </PREFIX>
  <SOURCE>
    <ID>source</ID>
    <ENDPOINT>http://aemet.linkeddata.es/sparql</ENDPOINT>
    <VAR>?a</VAR>
    <PAGESIZE>-1</PAGESIZE>
    <RESTRICTION>?a rdf:type <http://www.w3.org/2003/01/geo/wgs84_pos#Point></RESTRICTION>
    <RESTRICTION>?a <http://www.w3.org/2003/01/geo/wgs84_pos#long> ?x</RESTRICTION>
    <RESTRICTION>?a <http://www.w3.org/2003/01/geo/wgs84_pos#lat> ?y</RESTRICTION>
    <PROPERTY>rdfs:label AS lowercase->regularAlphabet</PROPERTY>
    <PROPERTY>geo:lat AS number RENAME slat</PROPERTY>
    <PROPERTY>geo:long AS number RENAME slong</PROPERTY>
  </SOURCE>
  <TARGET>
    <ID>target</ID>
    <ENDPOINT>http://linkedgeodata.org/sparql</ENDPOINT>
    <VAR>?b</VAR>
    <PAGESIZE>-1</PAGESIZE>
    <RESTRICTION>?b rdf:type <http://linkedgeodata.org/ontology/Windsock></RESTRICTION>
    <RESTRICTION>?b <http://www.w3.org/2003/01/geo/wgs84_pos#long> ?x</RESTRICTION>
    <RESTRICTION>?b <http://www.w3.org/2003/01/geo/wgs84_pos#lat> ?y</RESTRICTION>
    <PROPERTY>rdfs:label AS lowercase->regularAlphabet</PROPERTY>
    <PROPERTY>geo:lat AS number RENAME tlat</PROPERTY>
    <PROPERTY>geo:long AS number RENAME tlong</PROPERTY>
  </TARGET>
  <METRIC>AND(levenshtein(a.rdfs:label,b.rdfs:label)|0.8, euclidean(a.slat|slong,b.tlat|tlong)|0.8)</METRIC>
  <ACCEPTANCE>
    <THRESHOLD>0.8</THRESHOLD>
    <FILE>accept.nt</FILE>
    <RELATION>owl:sameAs</RELATION>
  </ACCEPTANCE>
  <REVIEW>
    <THRESHOLD>0.7</THRESHOLD>
    <FILE>review.nt</FILE>
    <RELATION>owl:sameAs</RELATION>
  </REVIEW>
  <EXECUTION>Simple</EXECUTION>
  <OUTPUT>NT</OUTPUT>
</LIMES>
```

On-the-fly recommender

GeoLOD

DATASETS

CLASSES

RECOMMENDER

1. Define a spatial dataset (point geometries)

☒ SPARQL Endpoint ☐ Shapefile ☐ GeoJSON

Get Classes

*Enter a valid SPARQL Endpoint e.g. <https://dbpedia.org/sparql>.
Press Get Classes to retrieve classes that contain 10 to 10000
spatial point entities using the W3C Basic Geo ontology.

Parse

2. Dataset Details



Airports

Entities: 37

Get Recommendations *

* It may take several minutes depending on the dataset's size and extent

Recommendations

10 Showing 1 to 10 of 15 classes

1 2 >

Label	URI	Dataset	Est. Related Instances	Rank	Link Discovery
Enter Label...		Enter Dataset...			
AirportsInGreece	http://dbpedia.org/class/yago/AirportsInGreece	DBpedia in Greek	4	1	☐ S ☐ L
Airport_Physical	http://umbel.org/umbel/rc/Airport_Physical	DBpedia in Greek	4	2	☐ S ☐ L
Park	http://linkedgeodata.org/ontology/Park	LinkLion - A Link Repository for the Web of Data	2	3	☐ S ☐ L
Airport	http://linkedgeodata.org/ontology/Airport	LinkedGeoData	2	4	☐ S ☐ L
GroundWaterBodyDeclaration	http://rdfdata.eionet.europa.eu/wise/ontology/GroundWaterBodyDeclaration	European Pollutant Release and Transfer Register (E-PRTR)	9	5	☐ S ☐ L
Airport	http://linkedgeodata.org/ontology/Airport	LinkLion - A Link Repository for the Web of Data	2	6	☐ S ☐ L
Pitch	http://linkedgeodata.org/ontology/Pitch	LinkLion - A Link Repository	2	7	☐ S ☐ L

REST API

GeoLOD services (the left part of the Request URI is <http://snf-661343.vm.oceanos.grnet.gr>).

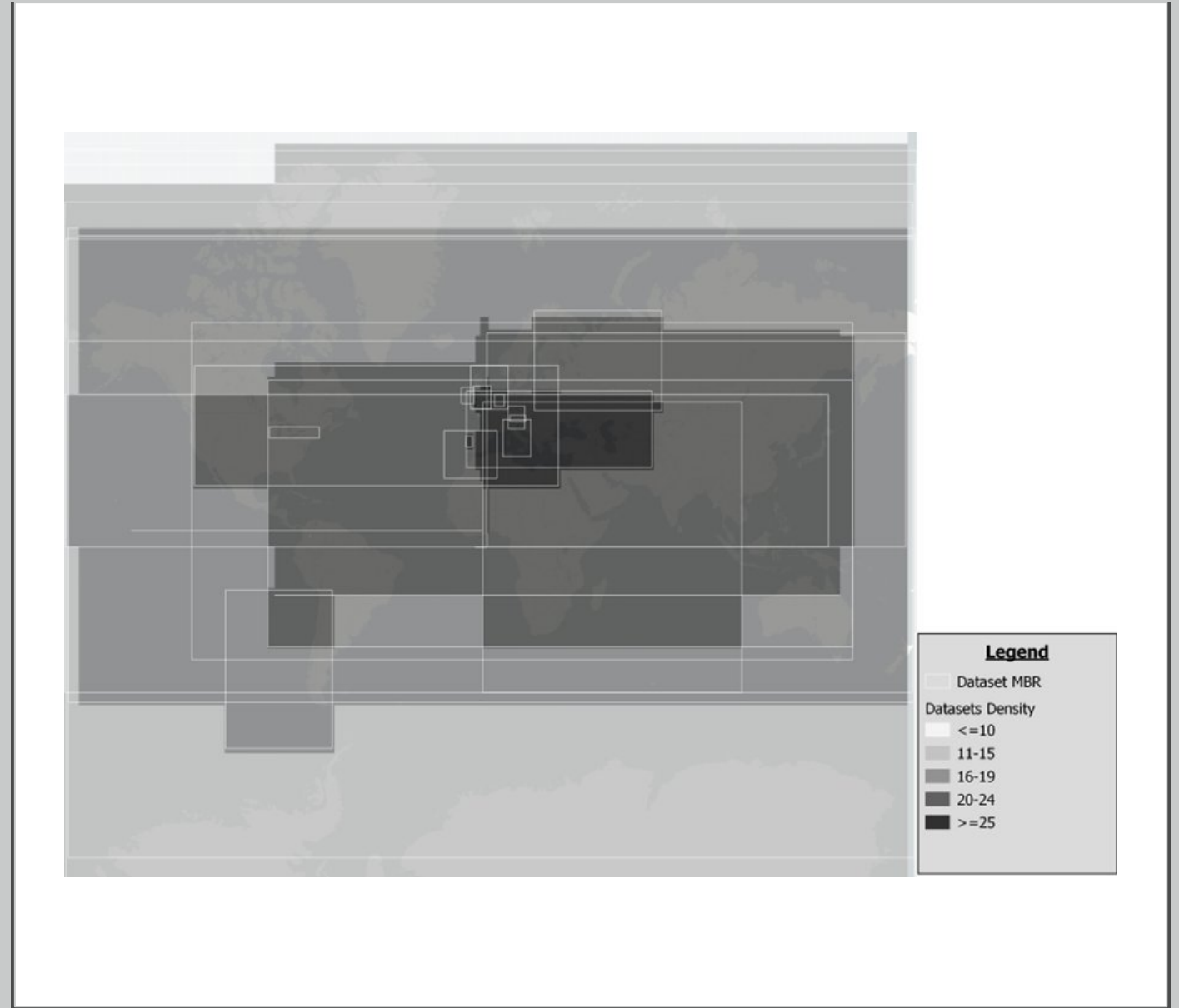
Service Name	Request URI	Description
GeoLOD Description	/api/download/dcat	Returns a DCAT-compliant turtle file that contains general information about GeoLOD and the list of the datasets in the Catalog
Dataset List	/api/datasets	Returns, in JSON, the list of datasets with their metadata (including internal dataset IDs) in the GeoLOD Catalog
Dataset Description	/api/datasets/<ID>	Returns, in JSON, the specified dataset metadata with the list of its classes. The dataset ID is a variable corresponding to the internal dataset ID. (e.g., http://snf-661343.vm.oceanos.grnet.gr/api/datasets/915 returns the metadata for the <i>AEMET</i> dataset)
Class List	/api/classes	Returns, in JSON, the list of classes with their metadata (including internal classes IDs) in the GeoLOD Catalog.
Class Description	/api/classes/<ID>	Returns, in JSON, the specified class metadata with the list of its recommended classes. The class ID is a variable corresponding to the internal class ID. (e.g., http://snf-661343.vm.oceanos.grnet.gr/api/classes/139090 returns the metadata for the <i>CaveEntrance</i> class of <i>Linklion</i> dataset).
Dataset GeoVoid	/api/download/geovoid/<ID>	Returns, in turtle format, the GeoVoid description of the specified dataset.
Dataset Recommendations	/api/download/datasetrecommendations/<ID>	Returns, in JSON, the list of recommendations for all specified dataset classes.
Class Recommendations	api/download/classesrecommendations/<ID>	Returns, in JSON, the list of recommendations for the specified class.

Implementation

- The GeoLOD backend API was developed in **Node.js** and the frontend application in **React**
- The GeoLOD database is a **PostgreSQL** with the **PostGIS** extension for spatial data management
- Queries to the SPARQL endpoints were sent with the **Fetch SPARQL endpoint** node.js module
- Bounding box thumbnails depicting were generated with the **Static Image Mapbox API**, and the interactive maps were built on **Leaflet** and **OpenLayers**
- GeoLOD is hosted in a Ubuntu 18 LTS 4GB Virtual Machine, provided by **oceanos**, a GRNET cloud Infrastructure as a Service (IaaS) for Greek academic institutes
- GeoLOD content, that is the list of spatial datasets and classes with their metadata and the recommendation lists for all classes, is **updated automatically every two months**, as a background process

Findings & Discussion

- GeoLOD contains 82 spatial datasets (out of 629 unique and 152 active SPARQL endpoints) and 5,218 spatial classes.
- They spatially cover all the world. Most datasets are “global”.
- In many datasets, georeferencing errors were found.
- GeoLOD provides approximately 90,000 class recommendations and on average it recommends 25 relevant classes per class.
- It requires on average 18 minutes to generate the class recommendation list for each class (although the execution time varies from a few seconds to several minutes), while LIMES would require approx. 5 hours for pair-wise class comparisons.



Pointers for Future Work

- Inclusion of datasets provided as RDF dumps and containing line and polygon geometries
- Extension of the on-the-fly recommender to support SPARQL endpoints that use additional spatial vocabularies (other than W3C Basic Geo) and additional spatial data formats, such as the Web Feature Service (WFS)
- Further improvement of the recommendation algorithm in terms of efficiency and effectiveness (e.g., integration with textual similarity)
- Integration with Silk or LIMES web services for the instant generation of instance link matching

Thank you!



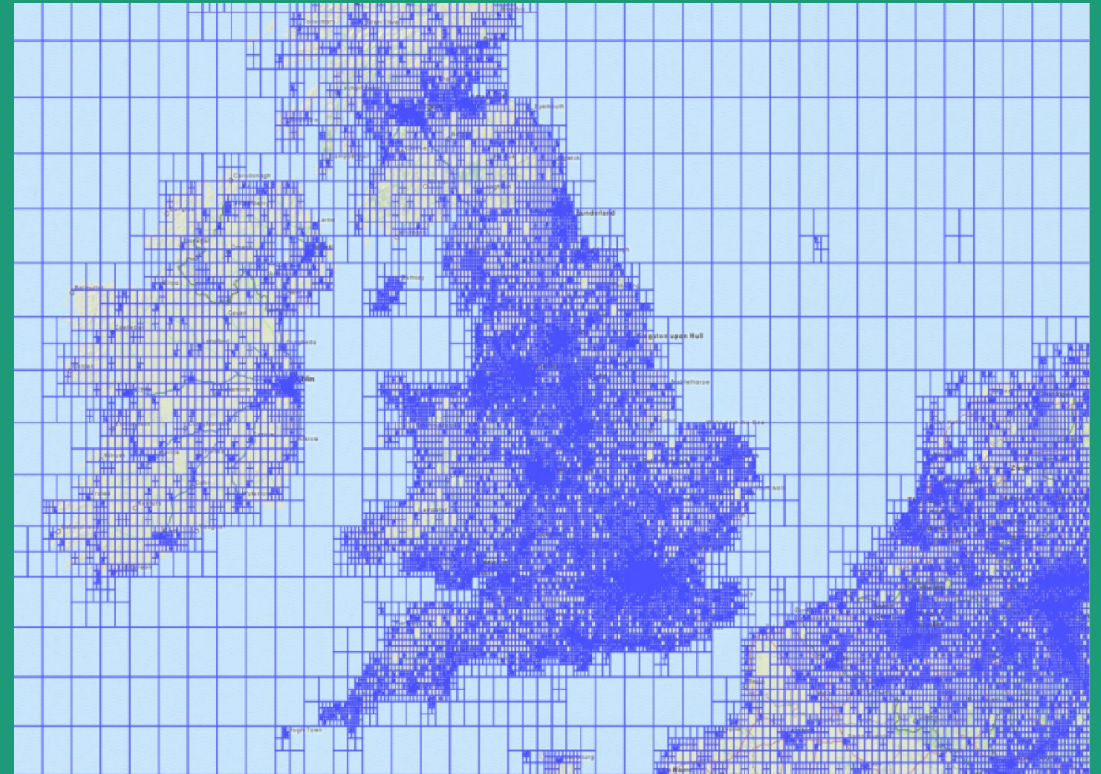
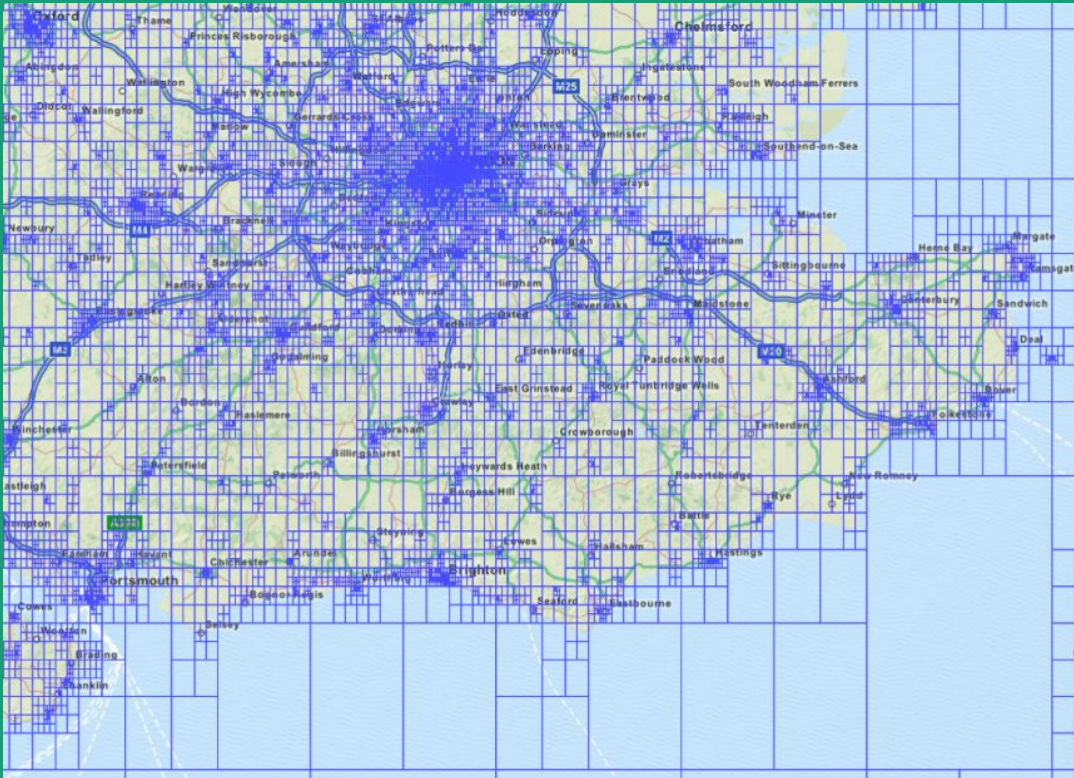
30 May 2022 in Hersonissos, Greece

SELECT Queries

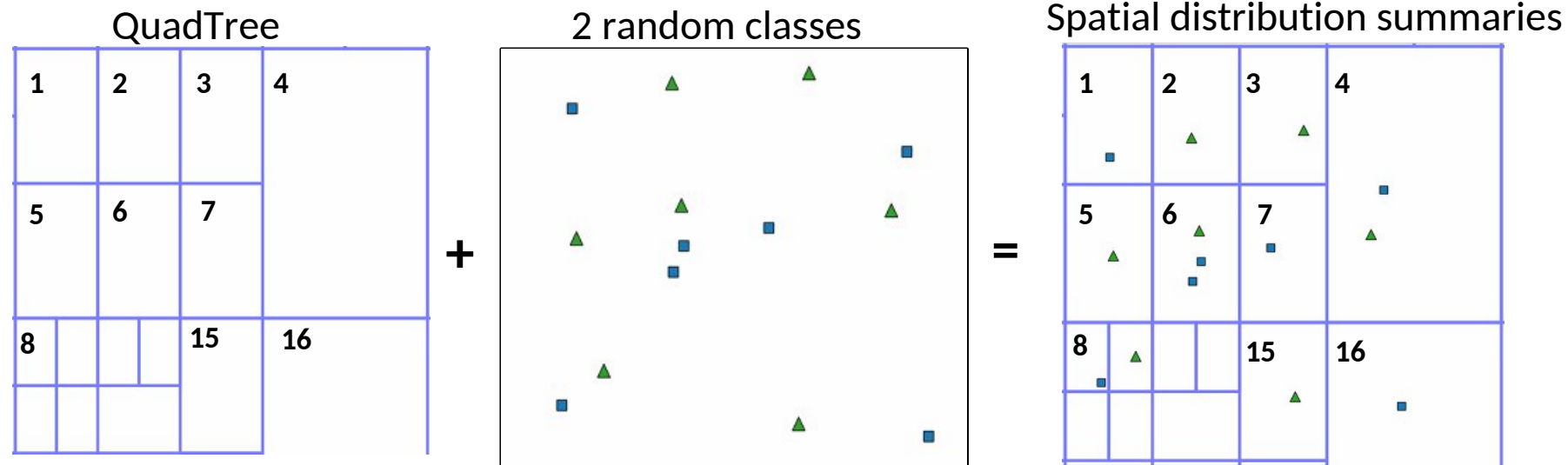
Spatial Vocabulary	SELECT Query
GeoVocab	<pre>SELECT DISTINCT ?class { ?geom <http://www.w3.org/2003/01/geo/wgs84_pos#long> ?x. ?geom <http://www.w3.org/2003/01/geo/wgs84_pos#lat> ?y. ?s <http://geovocab.org/geometry#geometry> ?geom. ?s <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> ?class}</pre>
GeoSPARQL	<pre>SELECT DISTINCT ?class { ?s <http://www.opengis.net/ont/geosparql#hasGeometry> ?geom. ?geom <http://www.opengis.net/ont/geosparql#asWKT> ?wkt. ?s <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> ?class}</pre>
GeoNames	<pre>SELECT DISTINCT ?class { ?s <http://www.w3.org/2003/01/geo/wgs84_pos#long> ?x. ?s <http://www.w3.org/2003/01/geo/wgs84_pos#lat> ?y. ?s <http://www.geonames.org/ontology#featureClass> ?class.}</pre>
W3C Basic Geo	<pre>SELECT DISTINCT ?class { ?s <http://www.w3.org/2003/01/geo/wgs84_pos#long> ?x . ?s <http://www.w3.org/2003/01/geo/wgs84_pos#lat> ?y. ?s <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> ?class.}</pre>
GeoRSS	<pre>SELECT DISTINCT ?class { ?s <http://www.georss.org/georss/point> ?point. ?s <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> ?class}</pre>

QuadTree

A global Quadtree index with not equally sized cells that capture global spatial instances distribution



Class Summarization



Summary of Squares class = {1,4,6,7,8,16}

Summary of Triangles class = {2,3,4,5,6,9,15}

.....

Class Relevancy

The more similar the summaries of classes the more similar the spatial distribution

Measurement of class summaries:

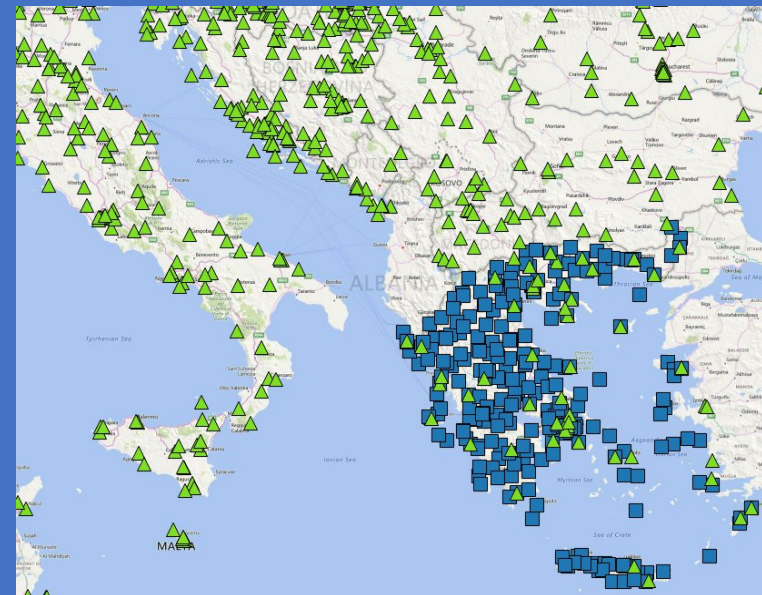
- Number of common cells
- Jaccard index
- Overlap Coefficient
- (Pointwise) Mutual Information
- Poisson Distribution
- ...

The algorithm returns a ranked list of relevant classes based on the values of the above metrics

Measurement of geographic similarity

Aim ☾ measure the similarity of classes spatial distribution.
Based on their summaries

*Similar spatial summaries ☾ similar spatial distribution ☾ classes
with related instances*

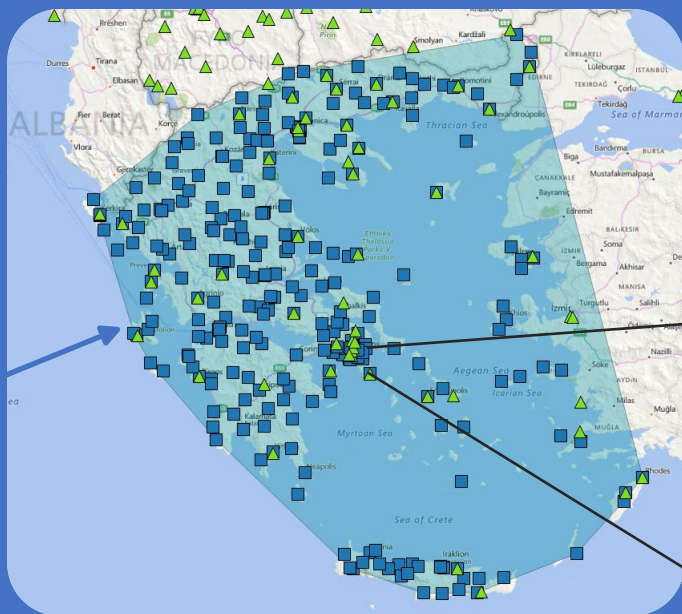


□ (S) <http://greek-lod.math.auth.gr/fire-brigade/resource/subdivisions> (Hellenic Fire Brigade)

▲ (T) <http://linkedgeo.org/ontology/FireStation> (LinkedGeoData)

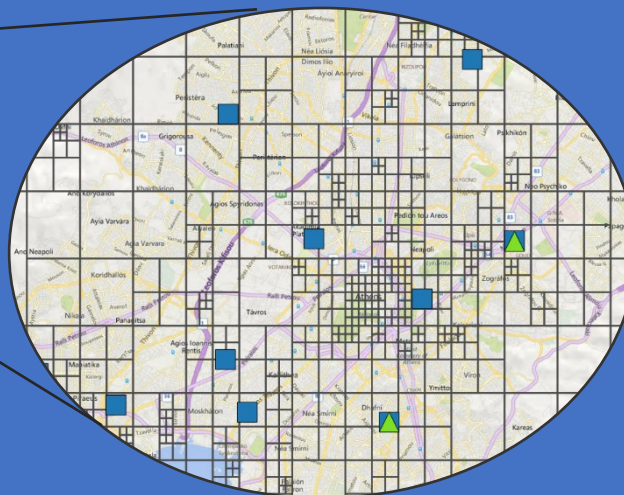
Μέτρηση γεωγραφικής ομοιότητας

- Κλάσεις που εκτείνονται σε διαφορετικές περιοχές (μη αλληλεπικαλυπτόμενο ConvexHull) δεν παρουσιάζουν όμοια γεωγραφική κατανομή και δεν είναι πιθανό να περιέχουν σχετικές οντότητες.
- Μέτρηση της γεωγραφικής ομοιότητας των συνόψεων 2 κλάσεων, S και T, στον αλληλεπικαλυπτόμενο χώρο τους (intersecting ConvexHull).



Intersecting
ConvexHull

- (S) <http://greek-lod.math.auth.gr/fire-brigade/resource/subdivisions> (Hellenic Fire Brigade)
- ▲ (T) <http://linkedgeoata.org/ontology/FireStation> (LinkedGeoData)



Υπολογισμοί από συνόψεις

$$\begin{aligned} |S| &= 275 & |T| &= 54 \\ |S \cap T| \text{ or } c &= 29 \\ |Q| &= 13.691 \end{aligned}$$

Spatial Similarity Metrics I

- **Number of Common Cells (C):** Ο αριθμός των στοιχείων της τομής των δύο συνόψεων.

$$C = |S \cap T|$$

Παρατήρηση: Δεν λαμβάνεται υπόψη το μέγεθος των συνόψεων.

- **Jaccard Similarity (JS):** Ο αριθμός της τομής (αριθμός κοινών κελιών) διά του αριθμού της ένωσης των δύο συνόψεων.

$$JS = \frac{|S \cap T|}{|S \cup T|}$$

Παρατήρηση: Δεν λαμβάνεται υπόψη η διαφορά του μέγεθους των συνόψεων.

- **Overlap Coefficient (OC):** Ο αριθμός της τομής (αριθμός κοινών κελιών) διά του αριθμού του μικρότερου συνόλου.

$$OC = \frac{|S \cap T|}{\min(|S|, |T|)}$$

Παρατήρηση: Δεν λαμβάνεται υπόψη η συχνότητα εμφάνισης των οντοτήτων στο χώρο μελέτης.

Υπολογισμοί από συνόψεις

$$|S| = 275 \quad |T| = 54$$

$$|S \cap T| \text{ or } c = 29$$

$$|S \cup T| = 300$$

$$|Q| = 13.691$$

Υπολογισμός δεικτών

$$JS = 0.09 \quad OC = 0.53$$

Spatial Similarity Metrics II

Η συχνότητα (άρα και πιθανότητα) εμφάνισης των οντοτήτων μίας κλάσης στο χώρο μελέτης υπολογίζεται ως το μέγεθος της σύνοψης της κλάσης διά τον αριθμό των συνολικών κελιών στο χώρο μελέτης.

$$P(s) = \frac{|S|}{|Q|}$$

Αν οι συνόψεις 2 κλάσεων είναι ανεξάρτητες, σύμφωνα με τη συνθήκη ανεξαρτησίας ενδεχομένων, ο αναμενόμενος αριθμός των κοινών κελιών τους είναι:

$$P(S \cap T) = P(S)P(T) \Rightarrow \hat{c} = \frac{|S||T|}{|Q|}$$

Αν ο πραγματικός αριθμός των κοινών κελιών των συνόψεων 2 κλάσεων (C) διαφέρει από τον αναμενόμενο αριθμό των κοινών κελιών των κλάσεων σε περίπτωση που ήταν ανεξάρτητες ($\hat{c}$), τότε οι κλάσεις σχετίζονται.

Ο λόγος συσχέτισης (R) μπορεί να υπολογιστεί ως ο αριθμός των κοινών κελιών διά τον αναμενόμενο αριθμό των κοινών κελιών στην περίπτωση που τα δύο σύνολα είναι ανεξάρτητα:

$$R = \frac{C}{\hat{c}}$$

- **Pointwise Mutual Information (PMI):** Ο λογάριθμός της πιθανότητας της ταυτόχρονης εμφάνισης δύο ενδεχομένων διά το γινόμενο των πιθανοτήτων τους:

$$PMI = \log R$$

Υπολογισμοί από συνόψεις

$$|S| = 275 \quad |T| = 54$$

$$|S \cap T| \text{ or } c = 29$$

$$|S \cup T| = 300$$

$$|Q| = 13.691$$

Υπολογισμοί Δεικτών

$$PMI = 1.42$$

Spatial Similarity Metrics III

- **Hypergeometric Distribution Culmulative Probability (HG):** Η πιθανότητα μία κλάση (S) να έχει c ή παραπάνω κοινά κελιά με μία κλάση (T) σε μία περιοχή που καλύπτεται από (Q) κελιά:

$$P(X \geq c) = \sum_c^{\min(|S|, |T|)} P(X = c) \text{ όπου } P(X = c) = \frac{\binom{|S|}{c} \binom{|Q|-|S|}{|T|-c}}{\binom{|Q|}{|T|}}$$

- **Poisson Distribution Culmulative Probability (PD):** Η πιθανότητα να εμφανίζονται c ή παραπάνω κελιά σε μία περιοχή που αναμένεται να εμφανίζονται $\lambda = \hat{c}$ κελιά:

$$P(X \geq x) = \sum_{x=|C|}^{\min(|S|, |T|)} P(x) \quad \text{όπου} \quad P(x) = \frac{e^{-\hat{c}} \hat{c}^x}{x!}$$

- **Δείκτης συσχέτισης δυαδικών μεταβλητών (Phi Coefficient).** Μία κλάση αναπαρίσταται από μία δυαδική μεταβλητή που παίρνει την τιμή 1 για τα κελιά που υπάρχουν στην σύνοψη μίας κλάσης και την τιμή 0 για τα κελιά που δεν υπάρχουν.

	T ₁	T ₀	Total
S ₁	C	S - C	S
S ₀	T - C	Q - S - T + C	Q - S = S'
Total	T	Q - T = T'	Q

2X2 πίνακας συνάφειας των κλάσεων S και T

$$\Phi = \frac{|C|*|Q|-|S|*|T|}{\sqrt{|S||S'|*|T||T'|}}$$

Υπολογισμοί από συνόψεις

$$|S| = 275 \quad |T| = 54$$

$$|S \cap T| \text{ or } c = 29$$

$$|S \cup T| = 300$$

$$|Q| = 13.691$$

Υπολογισμος δεικτών

$$HG = 2.37E-41$$

$$PD=0 \quad \Phi=0.23$$