

Web of Data Project. Exploring Spatial Knowledge with Knowledge Graphs. Analysis of European Capitals and UNESCO Heritage Sites

Khrystyna Ambroziak

Abstract

This project aims to identify the European capital with a population between 500,000 and 1 million that is within 50 km of the highest number of UNESCO heritage sites. The goal of this analysis is to provide insights into which capital city with a medium population is most suitable for visiting multiple UNESCO heritage sites within a short distance.

Motivation

The motivation behind this question stems from the desire to determine a travel destination in Europe that balances:

- A moderate population size, providing a less crowded yet urban experience.
- Proximity to a large number of UNESCO heritage sites, allowing visitors to explore culturally and historically significant locations efficiently.

This study combines geographic and demographic data to make an informed recommendation for cultural travelers.

Methodology

Data Collection

The project utilized data about:

- ****European capitals****: Names, geographic coordinates, populations, and country associations.
- ****UNESCO heritage sites****: Names and geographic coordinates.

The data was obtained using the following SPARQL queries via the Wiki-data Query Service:

Query 1: European Capitals

```

SELECT ?capitalLabel ?countryLabel ?coordinate ?
      capitalID
WHERE {
    ?country wdt:P31 wd:Q6256;    # Country
              wdt:P36 ?capital.    # Capital
    ?capital wdt:P625 ?coordinate; # Coordinates
              wdt:P1376 ?realCountry.
    ?country wdt:P30 wd:Q46.      # Europe
    SERVICE wikibase:label { bd:serviceParam wikibase:
      language "[AUTOLANGUAGE],en". }

    FILTER (?country = ?realCountry)

    # Extract Q-code of the capital
    BIND(STR(AFTER(STR(?capital), "http://www.wikidata.org
      /entity/Q") AS ?capitalID)
  }
ORDER BY ?capitalLabel

```

Query 2: UNESCO Heritage Sites

```

SELECT ?siteLabel ?coordinate ?siteID
WHERE {
    ?site wdt:P1435 wd:Q9259;    # UNESCO World Heritage
      Sites
              wdt:P625 ?coordinate. # Coordinates
    SERVICE wikibase:label { bd:serviceParam wikibase:
      language "[AUTOLANGUAGE],en". }

    # Extract numeric identifier

```

```

    BIND(STRATER(STR(?site), "http://www.wikidata.org/
        entity/Q") AS ?siteID)
}
ORDER BY ?siteLabel

```

The results of these queries were saved in the following files:

- `data/capitalLocations.csv` – containing data about European capitals.
- `data/unescoLocations.csv` – containing data about UNESCO heritage sites.

Convert to a Knowledge Graph

To enable advanced querying and geospatial analysis, the collected CSV data was transformed into a Knowledge Graph using Ontorefine. The transformation process included:

- Mapping CSV data to RDF using appropriate vocabularies, including GeoSPARQL for geospatial data representation.
- Creating RDF triples for capitals and UNESCO sites.

The resulting Knowledge Graph was imported into the GraphDB triple store. The following files were used for the import:

- `DataAfterOntotext/capital-result-triples.ttl` – RDF triples for European capitals.
- `DataAfterOntotext/unesco-triples.ttl` – RDF triples for UNESCO heritage sites.

Federated Query for Population Data

To enrich the local graph with population data for capitals with populations between 500,000 and 1 million, a federated SPARQL query was executed. The query retrieves population data from Wikidata and inserts it into the local graph.

Federated Query for Population Retrieval

```
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT DISTINCT ?capital ?label ?population
WHERE {
  SERVICE <https://query.wikidata.org/sparql> {
    ?country wdt:P30 wd:Q46 . # Europe
    ?capital wdt:P31 wd:Q515 ; # Instance of: capital
    wdt:P1082 ?population ; # Population
    wdt:P17 ?country .
    ?capital rdfs:label ?label .
    FILTER(LANG(?label) = "en")
    FILTER(?population > 500000 && ?population <
      1000000)
  }
}
LIMIT 100
```

Insert Query for Local Graph

```
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX wod: <https://paris-saclay.fr/project/wod/>

INSERT {
  ?capital wod:population ?population .
}
WHERE {
  SERVICE <https://query.wikidata.org/sparql> {
    ?wikidataCapital wdt:P31 wd:Q515 ;
    wdt:P1082 ?population ;
    wdt:P17 ?country .
    ?wikidataCapital rdfs:label ?label .
    FILTER(LANG(?label) = "en")
    ?country wdt:P30 wd:Q46 .
    FILTER(?population > 500000 && ?population <
      1000000)
  }
}
```

```

}
?capital a wod:Capital ;
          rdfs:label ?capitalName .
FILTER(LANG(?capitalName) = "en")
FILTER(STR(?capitalName) = STR(?label))
}

```

The queries and their results are stored in the following files:

- `FederatedQuery/federatedQuery.ttl` – SPARQL query for population retrieval.
- `FederatedQuery/insertFederatedQuery.ttl` – SPARQL query to insert population data into the graph.
- `Data/capital-with-population.csv` – capital data after adding population.

Data Processing

The following steps were performed:

1. ****Population Filtering****: Capitals with populations between 500,000 and 1 million were identified using a federated SPARQL query. The query retrieved population data from Wikidata and matched it with local graph data.
2. ****Distance Calculation****: The distance between each capital and nearby UNESCO heritage sites was calculated using the GeoSPARQL function `geof:distance`.
3. ****Site Count****: For each capital, the number of UNESCO sites within a 50 km radius was counted.

Final Query to Answer the Question

The final SPARQL query used to identify the capital city that meets the criteria is as follows:

```

PREFIX uom: <http://www.opengis.net/def/uom/OGC/1.0/>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>

```

```

PREFIX geof: <http://www.opengis.net/def/function/
    geosparql/>
PREFIX wod: <https://paris-saclay.fr/project/wod/>

SELECT ?capitalLabel ?capitalWKT ?population
    (GROUP_CONCAT(?siteLabel; separator=",-") AS ?
        siteLabels)
    (GROUP_CONCAT(?siteWKT; separator=",-") AS ?
        siteCoordinates)
    (COUNT(?site) AS ?numSites)
WHERE {
    # Capital cities in Europe
    ?capital a wod:Capital ;
        rdfs:label ?capitalLabel ;
        wod:population ?population;
        geo:hasGeometry ?capitalGeom .
    ?capitalGeom geo:asWKT ?capitalWKT .

    # UNESCO sites
    ?site a wod:Site ;
        rdfs:label ?siteLabel ;
        geo:hasGeometry ?siteGeom .
    ?siteGeom geo:asWKT ?siteWKT .

    # Bind the distance between the capital and the
    # UNESCO site
    BIND(geof:distance(?capitalWKT, ?siteWKT, uom:metre)
        AS ?distanceBetweenCapitalAndSite)

    # Filter for UNESCO sites within 50 km (50,000 meters
    # )
    FILTER(?distanceBetweenCapitalAndSite < 50000)
}
GROUP BY ?capitalLabel ?capitalWKT ?population
ORDER BY DESC(?numSites) DESC(?population)
LIMIT 1

```

This query is stored in the file:

- **FinalQuery/AnswerQuery.ttl** – containing the SPARQL query used to answer the main research question.

Results

The analysis identified **Stockholm, Sweden** as the European capital meeting the criteria:

- Population: 984,748 (within the specified range).
- UNESCO sites within 50 km: 7, including:
 - Birka and Hovgården
 - Drottningholm Palace
 - Skogskyrkogården
 - Chinese Pavilion at Drottningholm
 - Drottningholm Palace Theatre
 - Gardens of Drottningholm Palace
 - Royal Domain of Drottningholm

This result is stored in the file for using it in visualization:

- `ResultQuery/query-result.csv` – containing the answer of query.

Visualization

The results of the query were visualized using the Python library `folium` (for visualization, I initially tried to use YASGUI SPARQL, but there was a problem with data transfer from GraphDB that I couldn't fix, which made my repository unable to be used as an endpoint for YASGUI SPARQL). The visualization includes:

- The capital city, marked with a blue icon.
- Nearby UNESCO heritage sites, marked with green icons.
- Marker clustering to avoid overlapping for closely located sites.

The Python code used to create the map is stored in (to create a map, you need to run the code, enter the path to the data file, and as a result, a map file is created):

- `Visualization/map.py`.

The resulting interactive map is saved in:

- `MapResult/map_with_results.html`.

As shown in Figure 1, the visualization highlights the capital city and its nearby UNESCO heritage sites.

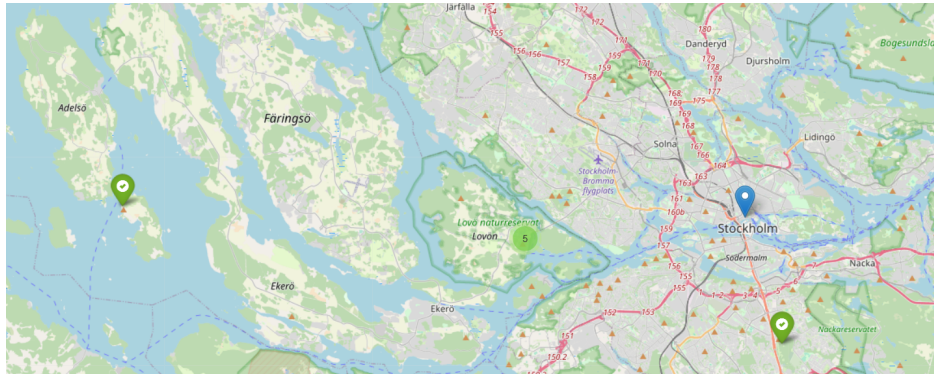


Figure 1: Visualization of Stockholm and nearby UNESCO heritage sites.

Conclusion

Stockholm stands out as the ideal travel destination for visitors interested in exploring UNESCO heritage sites. Its moderate population size offers a balanced experience between urban comfort and accessibility to culturally significant landmarks.

References

- Wikidata SPARQL Query Service: <https://query.wikidata.org/>
- UNESCO Heritage Sites Database: <https://whc.unesco.org/>
- Folium Documentation: <https://python-visualization.github.io/folium/>