SPARQL Lab Report

Timesheet

Exploration

Task	Time Spent (hh:mm)	Comments
Exploring predicates for continents	01:15	- Took some time to find the correct predicate for continents - Had to investigate sensible methods to remove sub continents and other entities which were classified as continent, but do not relate to the meaning of continent in Mondial.
Exploring predicates for countries	01:35	- Had many former countries, and unrecognised countries - Faced issues with namespaces and choice of predicates - Needed to use filters to get valid, current countries
Exploring predicates for cities	01:25	- Faced issues with namespaces and choice of predicates - Needed to get cities based on valid, current countries
Exploring predicates for provinces	02:50	- Could not retrieve provinces using the same predicates from cities and countries. - Had to further investigate how a province could be classified in dbpedia. - Some provinces included clerical regions, such as religious diocese, this required further investigation for removal.

Task	Time Spent (hh:mm)	Comments
Exploring predicates for organizations	01:20	- Dbpedia defines organizations differently to Mondial, organizations in dbpedia included schools and clubs - Research showed that organizations in Mondial are either political or economical organizations. This required further investigation on how these organizations can be specifically filtered.
Exploring predicates for languages	01:55	- Countries did not have one common property for spoken languages, this required further investigation of countries to identify the variations of how spoken languages are represented for each country.

Retrieval

Task	Time Spent (hh:mm)	Comments
Retrieving predicates for continents	00:35	- Most continents had predicates for Area. - Missing values for area was allowed, as I could not find a path to any other predicates which could represent area of continents.
Retrieving predicates for countries	01:25	- Countries did not use consistent predicates to respresent the same attribute. - Further work was needed to get possible candidate values for some attributes. - There were some missing values, in such instances a comprimise was made to either allow or fill-in a missing value.
Retrieving predicates for cities	02:00	- Faced some problems in retrieving a province for the city. - Some predicates were missing, e.g. population, latitude, longitude and elevation.

Task	Time Spent (hh:mm)	Comments
Retrieving predicates for provinces	03:00	- Most of the time was taken to find a way to get a capital city for provinces, escpecially where such values were missing, I had to decide on the best way to impute a value. - Took time to handle missing values for population and area. - Used existing filters from exploration to remove clerical/religious administrative regions.
Retrieving predicates for organizations	01:45	- Difficulty in retrieving formation/established dates, there wasn't a consistent way that organizations represented these values. - Information about the headquarter city/country/province were difficult to retrieve, most of it were missing and I did not have enough time to find alternative predicates that represent this information.
Retrieving predicates for languages	02:30	- Most of the time was spent on calculating the percentage of speakers. No langauges had a percentage of speakers, so it required a hacky way of calculating a percentage. - Some languages did not have a property for the number of speakers. I did not have enough time to find other properties which may refer to the number of speakers.

1 Continent

1.1 Exploration

```
SELECT DISTINCT ?continent
WHERE {
    ?continent rdf:type dbo:Continent.
    ?continent rdfs:label ?continentN.
    FILTER(lang(?continentN) = "en")
    BIND(REPLACE(?continentN, "@en", "" ) AS ?cName)
```

```
    FILTER(?cName in ("North America", "South America", "Oceania", "Asia", "Europe", "Africa"))
}
```

1.2 Retrieval

```
SELECT DISTINCT ?Name ?Area
WHERE {
    ?continent rdf:type dbo:Continent.
    ?continent rdfs:label ?continentN.
    FILTER(lang(?continentN) = "en")
    BIND(REPLACE(?continentN, "@en", "" ) AS ?Name)
    FILTER(?Name in ("North America", "South America", "Oceania", "Asia", "Europe", "Africa"))
    OPTIONAL{?continent dbo:areaTotal ?AreaTotal}
    BIND(?AreaTotal / 1000000 AS ?Area)
}
```

2 Country

2.1 Exploration

```
PREFIX dcterms: <http://purl.org/dc/terms/>
SELECT DISTINCT
    ?country
WHERE {
    ?country rdf:type dbo:Country.
    ?country dcterms:subject dbc:Member_states_of_the_United_Nations.
}
```

2.2 Retrieval

```
PREFIX dcterms: <http://purl.org/dc/terms/>
SELECT DISTINCT
    ?Name
    COALESCE(STR(?isoCode), ?Name) AS ?Code
    (SAMPLE(?capitalCity) AS ?Capital)
    (SAMPLE(?ProvinceValue) AS ?Province)
    (xsd:integer(COALESCE(?cPop, ?popEst, ?popCen)) AS ?Population)
    (xsd:integer(SAMPLE(?cArea) / 1000) AS ?Area)
WHERE {
    ?country rdf:type dbo:Country.
    ?country dcterms:subject dbc:Member_states_of_the_United_Nations.
    ?country rdfs:label ?countryN.
    OPTIONAL{?country dbo:iso31661Code ?isoCode.}
    FILTER(lang(?countryN) = "en")
    BIND(REPLACE(?countryN, "@en", "") AS ?Name)
    FILTER(?Name != "Member states of the United Nations")
    OPTIONAL{?country dbo:populationTotal ?cPop}
    OPTIONAL{?country dbp:populationEstimate ?popEst}
    OPTIONAL{?country dbp:populationCensus ?popCen}
```

```
    OPTIONAL{?country dbo:area ?cArea}
    {
        ?country dbo:capital ?capitalC.
    } UNION {
        ?country dbr:capital ?capitalC.
    }
    ?capitalC rdfs:label ?capitalCN.
    FILTER(lang(?capitalCN) = "en")
    BIND(REPLACE(?capitalCN, "@en", "") AS ?capitalCity)
    OPTIONAL{
        ?capitalC dbo:subdivision ?capitalSubdivisionEntity.
        ?capitalSubdivisionEntity rdfs:label ?capitalSubdivisionN.
        FILTER(lang(?capitalSubdivisionN) = "en")
        BIND(REPLACE(?capitalSubdivisionN, "@en", "") AS ?capitalSubdivision)}
    BIND(COALESCE(?capitalSubdivision, ?Name) AS ?ProvinceValue)
}
GROUP BY ?Name ?isoCode ?cPop ?popEst ?popCen
```

3 City

3.1 Exploration

```
PREFIX dcterms: <http://purl.org/dc/terms/>
SELECT DISTINCT ?country ?city
WHERE {
    ?country rdf:type dbo:Country;
        dcterms:subject dbc:Member_states_of_the_United_Nations.
    ?city rdf:type dbo:City;
        dbo:country ?country.
}
```

3.2 Retrieval

```
PREFIX dcterms: <http://purl.org/dc/terms/>
SELECT DISTINCT ?Name ?Country SAMPLE(?ProvinceValue) AS ?Province
?Population SAMPLE(?Lat) AS ?Latitude
SAMPLE(?Long) AS ?Longitude SAMPLE(?Elev) AS ?Elevation
WHERE {
    ?countryN rdf:type dbo:Country;
        dcterms:subject dbc:Member_states_of_the_United_Nations;
        rdfs:label ?countryLbl.
    OPTIONAL{?countryN dbo:iso31661Code ?isoCode.}
    ?cityN rdf:type dbo:City;
        dbo:country ?countryN;
        rdfs:label ?cityLbl.
    FILTER(lang(?countryLbl) = "en" && lang(?cityLbl) = "en")
    BIND(COALESCE(STR(?isoCode), STR(?countryLbl)) AS ?Country)
    BIND(STR(?cityLbl) AS ?Name)
    OPTIONAL{
```

```
        ?cityN dbo:subdivision ?subDiv.
        ?subDiv rdfs:label ?subDivLbl .
        FILTER(lang(?subDivLbl) = "en")
        BIND(STR(?subDivLbl) AS ?citySubDiv)}
    BIND(COALESCE(?citySubDiv, ?Country) AS ?ProvinceValue)
    OPTIONAL{?cityN geo:lat ?Lat}
    OPTIONAL{?cityN geo:long ?Long}
    OPTIONAL{?cityN dbp:elevationM ?Elev}
    OPTIONAL{?cityN dbp:populationTotal ?Population}
}
GROUP BY ?Country ?Name ?Population
```

4 Province

4.1 Exploration

```
PREFIX dcterms: <http://purl.org/dc/terms/>
SELECT DISTINCT ?Province
WHERE {
    ?Province a dbo:AdministrativeRegion .
    FILTER NOT EXISTS {
        ?Province a dbo:ClericalAdministrativeRegion .
    }
}
```

4.2 Retrieval

```
PREFIX dcterms: <http://purl.org/dc/terms/>
SELECT DISTINCT ?Name ?Country xsd:integer(?population) AS ?Population
    xsd:integer(SAMPLE(?totalArea) / 1000) AS ?Area ?Capital ?Name AS ?CapProv
WHERE {
    ?countryP rdf:type dbo:Country;
        dcterms:subject dbc:Member_states_of_the_United_Nations;
        rdfs:label ?countryName.
    OPTIONAL{?countryP dbo:iso31661Code ?isoCode}
    ?region a dbo:AdministrativeRegion;
        dbo:country ?countryP;
        rdfs:label ?regionName.
    FILTER NOT EXISTS {
        ?region a dbo:ClericalAdministrativeRegion.
    }
    OPTIONAL{?region dbo:areaTotal ?totalArea}
    OPTIONAL{?region dbo:populationTotal ?population}
    OPTIONAL{?region dbo:capital ?capitalCity.
        ?capitalCity rdfs:label ?capitalLabel.
        FILTER(lang(?capitalLabel) = "en")}
    OPTIONAL{?region dbo:largestCity ?largestCity.
        ?largestCity rdfs:label ?largestCityLabel.
        FILTER(lang(?largestCityLabel) = "en")}
```

```
    FILTER(lang(?countryName) = "en" && lang(?regionName) = "en")
    BIND(COALESCE(STR(?isoCode), STR(?countryName)) AS ?Country)
    BIND(STR(?regionName) AS ?Name)
    BIND(COALESCE(STR(?capitalLabel), STR(?largestCityLabel)) AS ?Capital)
}
GROUP BY ?Name ?Country ?population ?Capital
```

5 Organization

5.1 Exploration

```
PREFIX dcterms: <http://purl.org/dc/terms/>
SELECT DISTINCT ?Organization
WHERE {
    ?Organization rdf:type dbo:Organisation;
        rdfs:label ?orgName;
        dcterms:subject ?category .
    FILTER(?category IN (
        dbc:Political_organizations, dbc:International_political_organizations,
        dbc:Economic_organizations, dbc:Trade_associations,
        dbc:Intergovernmental_organizations, dbc:International_economic_organizations,
        dbc:United_Nations))
    FILTER NOT EXISTS{
        ?Organization dcterms:subject ?excludeCategory.
        ?excludeCategory rdfs:label ?excludeLabel.
        FILTER(CONTAINS(LCASE(?excludeLabel), "school") ||
        CONTAINS(LCASE(?excludeLabel), "club"))
    }
}
```

5.2 Retrieval

```
PREFIX dcterms: <http://purl.org/dc/terms/>
SELECT DISTINCT ?Abbreviation ?Name
    SAMPLE(?HQCity) AS ?City SAMPLE(?HQCountry) AS ?Country
    SAMPLE(?HQProvince) AS ?Province SAMPLE(?Established) AS ?Established
WHERE {
    ?org rdf:type dbo:Organisation ;
        rdfs:label ?orgName .
    ?org dcterms:subject ?category .
    FILTER(?category IN (
        dbc:Political_organizations, dbc:International_political_organizations,
        dbc:Economic_organizations, dbc:Trade_associations,
        dbc:Intergovernmental_organizations, dbc:International_economic_organizations,
        dbc:United_Nations))
    FILTER NOT EXISTS{
        ?org dcterms:subject ?excludeCategory .
        ?excludeCategory rdfs:label ?excludeLabel .
```

```
        FILTER(CONTAINS(LCASE(?excludeLabel), "school") ||
            CONTAINS(LCASE(?excludeLabel), "club"))}
    OPTIONAL{?org dbp:abbreviation ?abbreviation.}
    OPTIONAL{?org dbp:nickname ?nickname.}
    OPTIONAL{?org dbp:establishedDate ?establishedDate. }
    OPTIONAL {?org dbp:formation ?formation .}
    OPTIONAL{?org dbo:foundingYear ?established.}
    OPTIONAL{?org dbo:headquarter ?hq .
        OPTIONAL{
            ?hq rdf:type dbo:City .
            ?hq rdfs:label ?hqLabel . FILTER(lang(?hqLabel) = "en")
            OPTIONAL{?hq dbo:country ?hqCountry .
                ?hqCountry rdfs:label ?hqCountryName .
                OPTIONAL{?hqCountry dbo:iso31661Code ?isoCode}
                FILTER(lang(?hqCountryName) = "en") }
            OPTIONAL{?hq dbo:subdivision ?hqRegion .
                ?hqRegion rdfs:label ?hqRegionName .
                FILTER(lang(?hqRegionName) = "en") }
            OPTIONAL{?hq rdfs:label ?hqCityName .
                FILTER(lang(?hqCityName) = "en") }}
        OPTIONAL {
            ?hq rdf:type dbo:Country .
            OPTIONAL{?hq dbo:iso31661Code ?isoCode}
            ?hq rdfs:label ?hqCountryName .
            FILTER(lang(?hqCountryName) = "en")}}
    FILTER(lang(?orgName) = "en")
    BIND(STR(?orgName) AS ?Name)
    BIND(COALESCE(STR(?abbreviation), STR(?nickname)) AS ?Abbreviation)
    BIND(COALESCE(STR(?hqCityName)) AS ?HQCity)
    BIND(COALESCE(STR(?hqRegionName)) AS ?HQProvince)
    BIND(COALESCE(STR(?isoCode), STR(?hqCountryName), STR(?hqLabel)) AS ?HQCountry)
    BIND(COALESCE(STR(?establishedDate), STR(?formation), STR(?established)) AS ?Established)
}
GROUP BY ?Name ?Abbreviation
```

6 Language

6.1 Exploration

```
PREFIX dcterms: <http://purl.org/dc/terms/>
SELECT DISTINCT ?country ?language
WHERE {
    ?country rdf:type dbo:Country ;
        dcterms:subject dbc:Member_states_of_the_United_Nations.
    {
        ?country dbo:language ?language .
    } UNION {
        ?country dbo:officialLanguage ?language .
    }
}
```

Some countries have a language (e.g. Malaysia) property whilst some have an officialLanguage (e.g. South Africa) property.

6.2 Retrieval

```
PREFIX dcterms: <http://purl.org/dc/terms/>
SELECT DISTINCT ?Country ?Language
    ((SAMPLE(?LSpeakers) / ?Population) * 100) AS ?Percentage
WHERE {
    ?country rdf:type dbo:Country ;
        dcterms:subject dbc:Member_states_of_the_United_Nations ;
        rdfs:label ?countryName .
    OPTIONAL{?country dbo:iso31661Code ?isoCode}
    OPTIONAL{?country dbo:populationTotal ?population.}
    {
        ?country dbo:language ?language .
        ?language rdfs:label ?languageLabel .
        OPTIONAL{?language dbp:speakers ?speakers.}
    } UNION {
        ?country dbo:officialLanguage ?language .
        ?language rdfs:label ?languageLabel .
        OPTIONAL{?language dbp:speakers ?speakers.}
    }
    FILTER(lang(?languageLabel) = "en")
    FILTER(lang(?countryName) = "en")
    FILTER(datatype(?speakers) = xsd:integer || datatype(?speakers) = xsd:decimal)
    BIND(COALESCE(STR(?isoCode), STR(?countryName)) AS ?Country)
    BIND(STR(?languageLabel) AS ?Language)
    BIND(xsd:integer(?population) AS ?Population)
    BIND(xsd:integer(?speakers) AS ?LSpeakers)
}
GROUP BY ?Country ?Language ?Population
```

7 Religion