● A description in your own words of the following steps of the assignment:

○The conversion of the datasets to RDF (including a description in your words of the instructions expressed in the RML mapping file) [​minimum 80 words​]

○The linking of the two datasets using LIMES or other methods [​minimum 80 words​]

●The number of RDF triples generated by applying the mapping file(s).

●The type and number of entities linked across the datasets.

●A list of the SPARQL queries used (clearly marking the question it answers) along with a short explanation of the rationale behind the design choices behind the query (maximum3 lines). ​Hint: ​there are multiple ways to formulate the same query, explain why you chose your formulation. Also, a large query often can be seen as composed of ‘sub-queries’. If you use such a query, explain what these sub-queries do and how their results help to answer the ultimate question.

Technical report on assignment 1 for Building & Mining Knowledge Graphs

1. Introduction

This document reports on the process of creating a knowledge graph when provided with two data files. To this end, it will first consider the conversion of the two data files of different types to RDF. After that, to combine the information retrieved from the different files, the graphs created by the Converted RDF Files will be linked. Finally, to test the working of the conversion and linking, several queries will be executed on the graph for which the answers are included in this report.

1. Conversion from the data files to RDF

* TO BE MENTIONED:
* Number of RDF Triples generated by applying the mapping file

1. Linking the RDF graphs

* TO BE MENTIONED:
* Type and number of entities linked across the datasets

1. Querying the combined graph

The following queries have been made:

1.List all countries with population less than 50,000 and order them from the smallest to the largest in terms of landmass area (square kilometres)

2.List the countries with the top 10 highest GDP values in 2017

3.List the countries with the top 10 highest ​increases​ in GDP between 1960 and 2017

4.For each continent, count the number of countries in that continent that are in the top 20 for highest ​increases​ in GDP between 1960 and 2017. Your answer should contain 2 values: 1) the continent and 2) the number of countries from the top 20 that are located in those continents (For example, Asia - 15, Europe - 5)

5.Construct the triples representing the GDP per capita for each country in 2017

6.Directly insert the triples representing the GDP per capita for each country in2017, into your triplestore on GraphDB in the graph namespace“​[http://kg-course/query​](http://kg-course/query%E2%80%8B)”

7.Bonus: feel free to propose original elaborated SPARQL queries getting new interesting insights from the dataset.

1. Conclusion

Finding working ontologies and setting up the database, mapping and linking systems turned out to be a substantial amount of work. But the database has working, commonly used ontologies and the answers to the queries are feasible responses to what was asked, so the project can be considered succesful