COVID-19 App

Big Data and Semantic Technologies Course

2019/2020

Developed by:

Antonino Durazzo

Francesco de Pertis

Table of Contents

1 Introduction

In this chapter, the purpose of the project and adopted technologies will be pointed out.

1.1 Purpose

The purpose of the project is building an application that allows a user to display some relevant statistics on the data about the current sanitary emergency, very well known as COVID-19. The idea is to build an ontology in order to give a semantics to open raw data, daily updated and freely distributed by the Department of Civil Protection afferent to the Government of the Italian Republic.
These data represent tests, swabs and cases of COVID-19 and refer to Italian Regions and the entire Italian Country.

1.2 Adopted Technologies

Several software technologies have been adopted to implement this project:

- **Protégé** - a free and open-source ontology editor and framework - was used to get a graphical view of our ontology schema.
- **Eclipse IDE** - an integrated development environment - was used to write code to implement the project.
- **Java JDK** was used to develop the different component of the project and the interfaces that let these components to work together.
- **Apache Jena** - a free and open source Java framework for building Semantic Web and Linked Data applications - was used to define, populate and interrogate our ontology.
- **Window Builder** - SWT and Swing Designer plug-in for Eclipse - was used to develop the Graphical User Interface.
- **JFreeChart** - a Java library - was used to develop the charts integrated in the GUI.

1.3 Overview

The project was developed entirely in Java language and it is divided in six classes:

- BarChart.java - defines a Bar Chart Object used in Home and Comparison Tabs of the GUI.
- CovidApp.java - defines the queries and the whole GUI.
- dataRegion.java - contains the different datatype values for every region.
- OntModelCovid.java - defines the Ontology Schema and Population.
- parseCSV.java - imports and parses data from CSV dataset
- TimeSeriesChart.java - defines a Time Series Chart Object used in Trend Tab of the GUI.

2 Implementation

In this chapter, the implementation phase is pointed out.

2.1 Ontology Schema definition
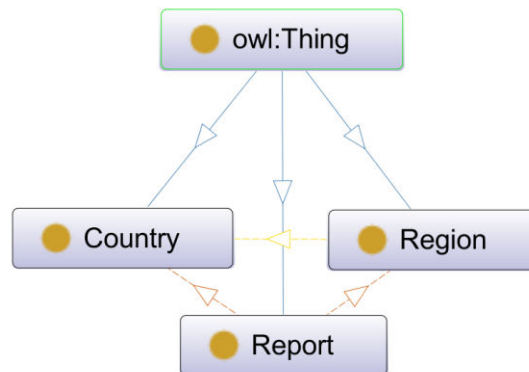
The ontology schema is defined as in the above figure. Three classes called Country, Region and Report are defined as sub-classes of owl:Thing, as requested in OWL praxis. The relations between these Classes are called Object Properties and they represent how these Classes are related to each other. In particular, Region "isPartOf" Country and a Report "belongsTo" only a Country or a Region.

Furthermore, these three sub-classes have some attribute that are called Datatype Proprieties. In particular:

- Country
  - hasName - property connected to a xsd:Name Literal
  - hasPopulation - property connected to a xsd:unsignedInt
  - hasLatitude - property connected to a xsd:double
  - hasLongitude - property connected to a xsd:double
- Region
  - hasName - property connected to a xsd:Name Literal
  - hasPopulation - property connected to a xsd:unsignedInt
  - hasCodRegion - property connected to a xsd:unsignedInt
  - hasLatitude - property connected to a xsd:double
  - hasLongitude - property connected to a xsd:double

- Report
  - o hasDate - property connected to a xsd:date
  - o hasDeceased - property connected to a xsd:unsignedInt
  - o hasDischargedHealed - property connected to a xsd:unsignedInt
  - o hasHomeIsolation - property connected to a xsd:unsignedInt
  - o hasIntensiveCare - property connected to a xsd:unsignedInt
  - o hasRecoveredWithSymptoms - property connected to a xsd:unsignedInt
  - o hasSwabs - property connected to a xsd:unsignedInt
  - o hasTestedCases - property connected to a xsd:unsignedInt
  - o hasTotalCases - property connected to a xsd:unsignedInt
  - o hasTotalHospitalized - property connected a xsd:unsignedInt
  - o hasTotalPositves - property connected to a xsd:unsignedInt

2.2 Ontology Population

Ontology Population consists of creating a sequence of individuals for every class. Country and Region population was made manually due to their small amount. In fact, in the ontology there is only one individual of the Class Country called "Italy" because the dataset under consideration regards only Italian Regions. Furthermore, there are twenty-one individuals of the Class Region that represent nineteen Italian Region and the only two Autonomous provinces of Trento and Bolzano.

Report population was made in an automatic way taking in input the dataset under consideration available to the following link:

https://raw.githubusercontent.com/pcm-dpc/COVID-19/master/dati-regioni/dpc-covid19-ita-regioni.csv

This dataset was parse with the methods defined in parseCSV.java Class. In particular, the program builds a new Report Individual for every row of the dataset creating a number of individuals equal to **number_of_regions** * **number_of_days**.

2.3 Ontology Interrogation and Graphical User Interface

Ontology Interrogation consists of different queries that are sent to the semantic dataset in order to get a particular datum of interest. In the project, the interrogation allows the user to view some important statistics about the data collected by the ontology. Queries let the user get one or more reports and view their datatype properties.

Graphical User Interface, abbreviated GUI, represents a very simple and easy way which the user can send queries to the semantic dataset and view their results. The GUI is developed with Java Swing API and it is divided in three tabs called Home, Trend and Comparison. The query sent to the semantic database is:

PREFIX NS: <http://www.covidapp.org/ontology#>

SELECT *

WHERE { ?Report NS:hasDate "**date**"^^<http://www.w3.org/2001/XMLSchema#date> .

?Report NS:belongsTo NS:**Region** .};

Home Tab allows the user to view the Datatype properties associated to a single Report that can be identified by a place and a date. By clicking on the Search Button, a Bar Chart will appear in another window. The query sent to the semantic database is:

PREFIX NS: <http://www.covidapp.org/ontology#>

SELECT *

WHERE { ?Report NS:belongsTo NS:**Region** .

?Report NS:hasDate ?**date** . }

ORDER BY ASC(?**date**) "

Trend Tab allows the user to view a Time Trend for a selected Region or the whole Country for a selected Datatype Property. By clicking on the Search Button, a Time Series Chart will appear in another window.

Comparison Tab allows the user to view a comparison between all the Regions and the whole Country a selected Datatype Property in the last possible date. The query sent to the semantic database is:

PREFIX NS: <http://www.covidapp.org/ontology#>

SELECT *

WHERE { ?Report NS:hasDate **"date"**^^<http://www.w3.org/2001/XMLSchema#date>.

?Report NS:**hasProperty** ?value . }" +

ORDER BY DESC(?value);

3) Conclusions