

# Techniques for Large Scale Data - Group 23

Julia Szulc - juliasz@student.chalmers.se  
Ahmed Groshar - gusgroah@student.gu.se

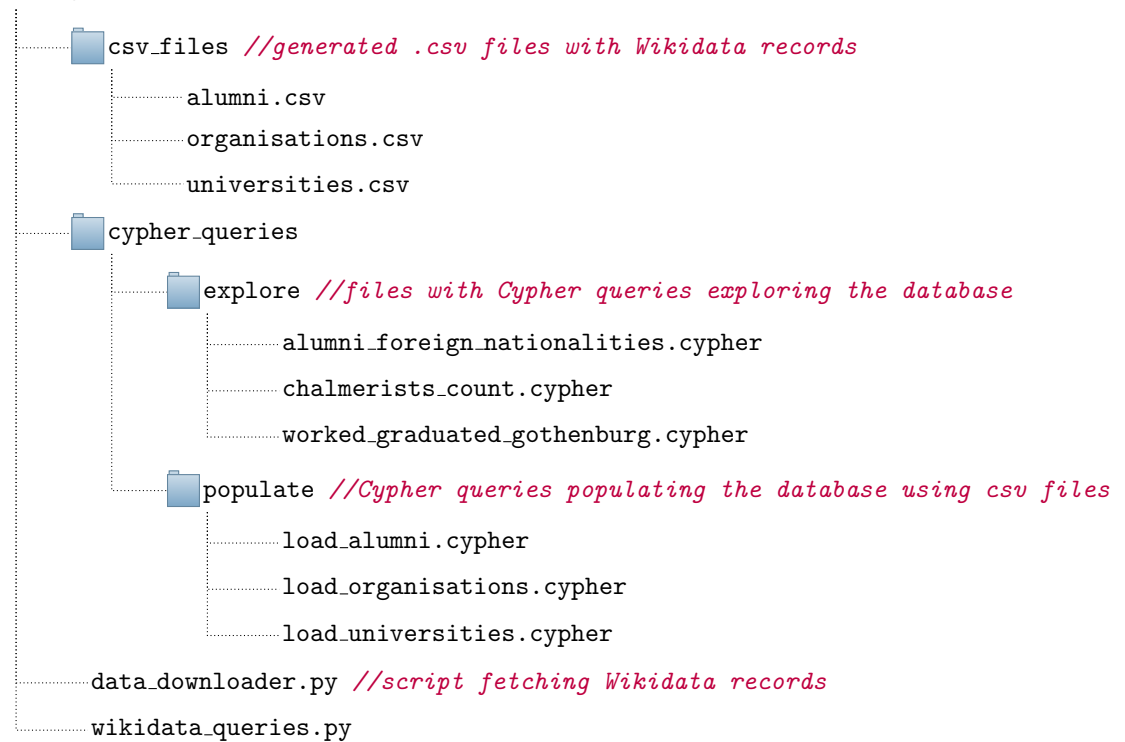
May 25, 2020

## Assignment 5

### 1

#### Directories and files

Assignment5



## Description of the solution

The records were obtained from Wikidata database using SPARQL scripts (see `wikidata_queries.py`) and saved to separate `.csv` files. All these actions were performed using a Python script (`data_downloader.py`). It had been assumed that the information provided by Wikidata are correct and that all requested data is available. The only exception was Place class as it had been noticed that many fetched organisations do not have any information about their locations assigned, as well as there are some alumni with unknown places of birth. In such situations these connections were simply not provided to the data points.

The `.csv` files were later uploaded to a GitHub repository and the direct URLs were passed to Cypher scripts populating the Neo4j database (see files in `cypher_queries/populate` directory). The queries were entered manually in a following order (to make sure that all dependencies between the nodes were valid):

1. `load_universities.cypher`
2. `load_alumni.cypher`
3. `load_organisations.cypher`

After introducing all the records to the database the general graph was inspected to check the correctness of created nodes and relationships between them. Additionally, a few Cypher queries exploring the database (see files in `cypher_queries/explore` directory) were tested.

## Question 2

### 2.1

The study case given is of a person that tests positive for COVID-19 that uses the public transport system frequently. That individual realizes he has access to information about which other passengers have travelled recently on the same buses and trams as the individual, and he has their contact details.

Public transport organization electronic ticketing options:

- travel card(registered in a system, together with the passenger's telephone number)
- mobile app

Regardless of the method of payment the platform must provide contact details especially during a pandemic as the potential harm can be life threatening for many if the not all members at least of a local society.

## Possible actions

If an individual that has tested positive for COVID-19 primarily he is to do one of the following:

1. Contact directly the transport organization
  - This option has the benefit that a third party person (being a professional that would professionally handle such a critical situation and perhaps give a more calm and accurate instructions to passengers that might me at risk) would be involved into the communication of the infected person to the people using certain bus or tram lines.
  - On the other hand this communication might take more time hence putting someone at risk.
2. Contact passengers directly
  - Perhaps faster communication between infected person and people using the transport, and in case there is no operator monitoring this can alert people directly.
  - On the other hand, this case is prone to misinformation or "fake news" or the infected person might not use the tool for signaling properly due to panic. This on its own might cause panic in certain lines and create more problems and worsen the situation.
3. Does not contact anyone
  - This is most probably the worst case, because there might be risk for some or many of the passengers.
  - This can be done consciously or unintentionally, that's why it can be a good advice of the medical staff to tell newly tested positive for COVID-19 patient to alert the travel organization in case of recent or frequent travel.

## Identify stakeholders

Below we list the mainly affected key groups of people that are relevant to the situation.

Potential stakeholders:

1. Transport organization
  - Administrative staff
  - Drivers

It would be beneficial for the transport organization if it is alert during the time of a pandemic by having more people ready to take calls from passengers signaling that they have tested positive for COVID-19 or that

have questions or concerns. The display of relevant and accurate information onto media where they display anything related to the organization can help the general public and increase public trust and encourage people to still the organization's transportation services.

A risk for the transportation organization could be a neglectful and undermined approach of the situation e.g. by not disinfection with the recommended chemicals. This could lead to a spread of the disease, undermined trust in the ability of the organization to deal with the situations and also reduced revenues from selling tickets.

## 2. Ticketing platform

- Tickets platform support personnel

It would be beneficial to the support personnel of the application if they make sure that there is clear information on the if there are individuals that have tested positive so that it can be appreciated as a fast and efficient way to connect the professional that could take a fast and measured decision on what measure can be taken and who has to be informed what, involving all stake holders.

The risk that could be that the platform may provide the means for someone to flood with fake information the operators responsible or if the application fails during critical times.

## 3. Travellers using the transport system:

- Healthy passengers(esp. vulnerable groups)
- Passengers that have just tested positive for COVID-19

What is beneficial for healthy passengers is to remain safe and to be able to conduct their commute safely. That could happen if they are told when it is or it is not safe to use the transport.

It is not beneficial to these stakeholders if they disregard advice from other key stakeholders.

## 4. Local health professionals

- Directors of hospitals and health care facilities
- Doctors, nurses, care workers

Local health professionals wish to not be overwhelmed with the number of patients making their work impossible or extremely difficult. So for them is beneficial to receive information of the number of infected people and expectations of infections.

It is risky for these stakeholders if they do not communicate with other key stakeholders from at least local government or in this case transport organization throughout a pandemic or at key moments or equally important if they are not told key information from other key stakeholders.

## 5. Local and Government authorities

- Mayors of cities and municipalities
- Government

This main group of various stakeholders that have power on different scale regard as beneficial if the spread of a disease is detected and infected individuals are under care unable to spread the disease, which bring the need for a lock-down at certain level. Functioning almost without interruption especially of critical infrastructure is vital for these stake holders as they are the main figures from which taking responsibility and action is expected

It is risky for these stakeholders if they do not assess the situation well taking into account all stakeholders and taking into consideration the professional advice into their policy making strategies.

## 6. Wider public and the world population

- Regional resident and national public
- International public

Residents locally even though not directly using the tram or bus lines will benefit from fast action and potential prevention of the spread of the disease via interlines travellers, namely that may transfer the disease from one line to completely different one. The same is valid for travels consequently that extend internationally.

The risk for these huge groups of potentially affected stakeholders can be huge as it becomes harder to track people overseas for instance, where different procedures might apply and different time frames might be needed for appropriate response to a threat of spreading.

## Conclusions

Taking all stakeholders into account, the goals of all might be viewed differently and the measures that have to be taken and the degree of "lock-down" if needed would be different among the stakeholders. All stakeholder's safety and concerns can be useful in crafting a plan for limiting the spread of such a disease. Regarding informing a group of people about potential risks and concerns would be most probably better done by professionals that can inform the general public in a more coordinated, precise clear manner, minimizing the risk of havoc. Furthermore, timely communication, consensus and abiding by the rules of all stakeholders are crucial for going through such a challenge.