

Building & Mining Knowledge Graphs

Individual Project

Building knowledge graph for exploring dependencies between
suicide, climate change and economics

Maastricht University
Faculty of Science and Engineering
Department of Data Science and Knowledge Engineering

Author

Martyna Mikos, i6155316

1 Introduction

Knowledge graphs (KG) play an important role in answering questions based on information from various sources like texts and databases. Reason for that being the structured knowledge format that allows for efficient querying (Moschitti et al., 2017). Knowledge graphs are without a doubt a valuable tool for uncovering interesting facts about the gathered data.

The goal of this project is creating a KG by interlinking 3 different datasets, which contain information about number of suicides, weather and economic data per country over the period of 37 years.

2 Problem statement

In this project the focus will be put on converting multiple datasets from different sources to RDF and integrating them in order to answer the following questions:

1. Does weather influence the suicide rate?
2. What other economic/demographic factors can influence that rate?
3. Are there any correlations and dependencies found in the data?
4. Based on the results can we draw any conclusions on how to create awareness and identify risks associated with the suicide rate studied?

3 Significance

It is important to identify and asses suicide risks and ways of protection in order to succeed in suicide prevention. Protective factors and risk identification will be a huge asset in providing changes and public awareness campaigns in order to decrease suicide rates (Center and Rodgers, 2011).

4 Related work

There is a lot of existing work on the topic. Current study shows a strong link between cold climate and suicide, which is stronger than the influence of the economic factors. Nonetheless both are proven to be existent (Fountoulakis et al., 2016).

The methods used in this project are quite commonly used in research papers regarding building knowledge graphs and combining multiple datasets. Usage is spread though various fields like developing KG as a tool for integrating multiple sources of medical data and analysis of health care quality (Huang et al., 2017). Further research shows that KG are also used for analysing enterprise information from many different websites and databases in order to achieve completeness and valuable investment analysis and data visualization (Ruan et al., 2016).

5 Methodology

In order to answer the research questions multiple datasets from different sources will be converted to RDF and integrated.

First dataset includes data for the time period between 1979 and 2016. It covers numbers by country, age and sex. The source of this data is WHO Mortality Database (WHO, 1979-2016).

Second dataset was retrieved from the World Bank (The World Bank, 1979-2016) and covers the same period of time as the WHO dataset. The economic indicators that will be used for the analysis are:

- Educational attainment, at least completed upper secondary
- Unemployment
- Poverty headcount ratio at national poverty lines
- Net migration
- Population
- GNI per capita
- GINI index

The third dataset is the weather data that shows the average temperature for each country. Data was retrieved from Berkeley Earth (Berkeley Earth, 1979-2016).

Based on the produced output questions about the created knowledge graphs will be answered. Tools that will be used for the graph exploration are SPARQL queries using GraphDB. Moreover RStudio will be used to write a script for performing statistical tests, analysis and visualization.

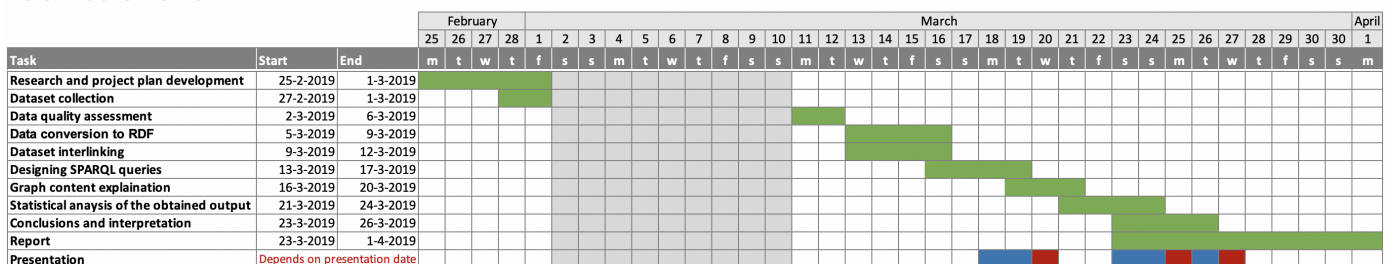
6 Milestones & deliverables

Difficulties in this project arise due to the problem of integrating multiple datasets in terms of mapping and URI. Moreover some other challenges might arise on the way that will have to be confronted. One of these being the data heterogeneity and achieving the integration and interoperability between the knowledge resources. There could also be a problem with missing data and inconsistency of the datasets that are discovered along the way.

Deliverables in this project are the RDF file SPARQL queries along with the R script written to perform any further analysis. On top of that the report and presentation will be delivered.

6.1 Time Management

Gantt chart



7 Innovation

There hasn't been much work done on the topic in terms of using a knowledge graph as a tool. Although the topic has been analysed and discussed many times in the papers from economics and social perspective, the idea of using knowledge graphs to analyse the problem hasn't been done before.

References

- Berkeley Earth (1979-2016). *Global Average Land Temperature by Country*. data retrieved from The Berkeley Earth Database, <http://berkeleyearth.org/data/>.
- Center, Suicide Prevention Resource and P. Rodgers (2011). “Understanding risk and protective factors for suicide: A primer for preventing suicide”. In: *Newton, MA: Education Development Center, Inc.*
- Fountoulakis, Konstantinos N et al. (2016). “Relationship of suicide rates with climate and economic variables in Europe during 2000–2012”. In: *Annals of general psychiatry* 15.1, p. 19.
- Huang, Zhisheng et al. (2017). “Constructing knowledge graphs of depression”. In: *International Conference on Health Information Science*. Springer, pp. 149–161.
- Moschitti, Alessandro et al. (2017). “Question Answering and Knowledge Graphs”. In: *Exploiting Linked Data and Knowledge Graphs in Large Organisations*. Springer International Publishing, pp. 181–212. ISBN: 978-3-319-45654-6.
- Ruan, Tong et al. (2016). “Building and exploring an enterprise knowledge graph for investment analysis”. In: *International Semantic Web Conference*. Springer, pp. 418–436.
- WHO (1979-2016). *Suicide, crude death rate per 100 000, both sexes*. data retrieved from WHO Mortality Database, <http://apps.who.int/healthinfo/statistics/mortality/whodpms/>.
- The World Bank (1979-2016). *Economic indicators*. data retrieved from World Development Indicators, <https://data.worldbank.org/indicator>.