|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **MEMORANDUM** | | A picture containing graphical user interface  Description automatically generated | | |
|  |  |  |  |  |
| **To:** | Dr Foaad Khosmood | | | |
|  | foaad@calpoly.edu |  |  |  |
| **From:** | Chandler Jones |  |  |  |
|  | cjones87@calpoly.edu |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| **Date:** | 5/7/2021 |  |  |  |
| **Re:** | Lab 2 – *Encyklopedia Solidarności* | | | |

Overview

In this project, we will be tackling a real-life history challenge. A resource by the name of *Encyklopedia Solidarności* is available that primarily summarizes the lives of certain Polish Cold-War Opposition figures in a series of encyclopedia entries written entirely in Polish.

The goals of the assignment are to:

* Become familiar with doing NLP on a language you don’t know.
* Gain insight into the interconnections of the opposition figures using NLP tools and techniques
* Visualize the relationships in a graph (described below)
* Write about the insights gained (supported by numbers) and display the visualization in a 3-5 page report.

Visualization

In this project, we use a 2\_D knowledge graph visualization tool to explore our insights. The method involves using python, and its various packages, including matplotlib, nltk, spacy, and pandas, to generate novel insights about a language we do not speak!

This is exactly the type of information I am interested in generating for my long-term project. This lab has been a novel endeavor, and a pursuit I plan to continue.

* People: these are the most important nodes and how the articles are organized
* Locations: names of cities, possibly buildings and other locations
* Groups: names of organizations that people belonged to or were associated with

After these, you should explore a few other concepts such as imprisonment, exile, court cases or events -- which are necessarily associated with a date (also necessarily tied to one or more locations)

After downloading and parsing the contents of the *Encyklopedia Solidarności,* I formed a dictionary which included some relevant data for each person in the encyclopedia. A single example is included below in Figure 1. I chose to hone in on the regions linked at the bottom of the page first, as my experience with BeautifulSoup is still growing. Starting with just the basics allowed me to push forward and begin plotting some knowledge graphs. I figured completing as much of the lab as I could, would benefit me more than struggling any more than necessary with the parsing portion of the lab.

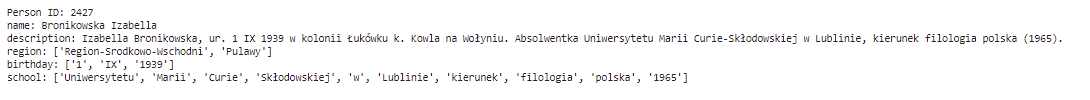


Figure 1. Sample Dictionary for each person

After forming this dictionary, the next step was following a guide for generating knowledge graphs. I was able to generate a Pandas dataframe with sources & targets for each person. My first goal was to plot where each person was from. The dataframe used is shown below in Figure 2.

A picture containing text, receipt, screenshot

Description automatically generated

Figure 2. Pandas Dataframe relating Individuals to their Birth Region

I then generated a directed graph showing the relations. As with the guide, this proved to be much too dense of a plot to gather any useful data.

A picture containing text, invertebrate, echinoderm

Description automatically generated

First Attempt using all regions as source, and names as target

I modified my dataframe, and excluded all regions except for a single randomly chosen region, Szczecin. This is a large city in Northwest Poland which I had not yet heard of. This graph was much more manageable, and data is beginning to emerge.

Chart, pie chart

Description automatically generated

Figure 2. Birth Regions using only regions corresponding to Szczecin

Continuing on this train of discovery, I then modified the data frame to a major city, the capital of Poland, Warsaw.

A picture containing text, receipt

Description automatically generated A picture containing text

Description automatically generated

Figure 3. Individuals by Birth Region, Warszawa

As can be seen, many more of the cold war opposition leaders were born in the capital. I hypothesize that the major cities included more military & political presence during this time period and thus had a larger impact on radicalizing the population against the cold war.

Using this train of thought, I began to be curious if in fact WWII, and the great depression had any telltale signs of inducing the opposition to ongoing political events. My first check was to assess the number of opposition leaders born during the Great depression. This is shown below in Figure 4.

Chart, pie chart

Description automatically generated

Figure 4. Opposition Leaders with Birthdays Corresponding to 1929

I had to figure out a way to compare the number of opposition leaders born by year, without plotting each and every one in a circular graph. I reconfigured my data and plotted the Members in Solidarnsci by year to get a better sense of when these people were born. The plot by year is shown below in Figure 5.

Chart, histogram

Description automatically generated

Figure 5. Members in Solidarnsci as they were born, year by year

A majorly uniform bell-curve can be seen above, and interestingly we see that there were a number of early adopters to the opposition, with a stark dropoff on the rightward tail. It seems that no individuals joined Solidarnsci (at least in this encyclopedia, after the year 1971.

After looking around at the timeline of the Cold War on Wikipedia, I hypothesize that as the war drew to a close, and significantly a treaty was signed on March 5th, 1970, there was no longer any major reason to join the opposition.

Graphical user interface, text, application

Description automatically generated

Figure 6. Excerpt of the 1970s section from “Timeline of the Cold War” Wikipedia Page

References:

<https://www.analyticsvidhya.com/blog/2019/10/how-to-build-knowledge-graph-text-using-spacy/>

Google Translate

<https://en.wikipedia.org/wiki/Szczecin>

<https://en.wikipedia.org/wiki/Timeline_of_events_in_the_Cold_War>