**Project Overview:**

The aim of this project is to analyse the language of the users of the /pol/ board on the popular imageboard 4chan.

1. **Creation of a database**: the website <https://archive.4plebs.org/pol/> provides a live-updated archive of all the threads posted on /pol/. Starting from the data collected in this archive, I want to create a database suitable to coding on python. Since the database is constantly updated, I will start from the first post that appears when I start creating the database and then include all the others published before that one. **Issue:** how to create a database deleting pictures, users’ codes and names and other useless information? In which format should I compile the database? Is the amount of data manageable?
2. **Text analysis**: using the database, I will first perform a word count, deleting all the “useless” words. Particular attention will be paid to hate speech, racist expressions, and “Internet words” and meme. Then, these words will be analysed in context in order to understand how they are used and how are some words being reshaped according to /pol/ users.
3. **Dictionary**: I will use the “special words” found by text analysis to compile a /pol/ dictionary.

**Project to be delivered:**

1. A dictionary of Internet words, memes and /pol/ language will be compiled and made accessible to everyone. Every entry of the dictionary will include the meaning of the word, its story and some related contents (source: <https://knowyourmeme.com/>).
2. Exactly as the dictionary, the database will be made available online.

**Project schedule:**

1. **Milestone 1:** creation of the database. (Deadline: end of November)
2. **Milestone 2**: understand how to use Phyton for my project. Writing the code for word count and context. (Deadline: end of December-start of January)
3. **Milestone 3:** writing a code to easily compile the dictionary. (End of January-second semester)
4. **Milestone 4:** compile the dictionary. (Second semester)
5. **Milestone 5:** find an online space to publish data and dictionary. (Second semester)