# Processing Language with Python: Final Projects

You will formulate a practical problem for which you program a tool or application in Python. Examples projects are given below, but we encourage you to think creatively about an issue/tool that is useful for yourself.

The final project applies all the skills and knowledge seen in the course but also goes beyond requiring you to make use of online resources and external libraries. The problem you intend to solve should be sufficiently extensive.

You are allowed to formulate a project in group or solo. When you choose a group project, each person will solve a well-defined part of the larger problem; each part should be sufficiently extensive per person.

***Deadlines:***

* Project proposal brainstorming post on Minerva forum: ***20 November 2017, 23:59h***.
* Finished project proposal with problem analysis: ***27 November 2017, 23:59h***.
* Final project submission: ***24 January 2018, 23:59h***.

The project accounts for 100% of the final grade. The graded assignments give you extra points on top of the project score.

## A. Project proposal:

### Step 1: Brainstorming

The first step in the project proposal is to explain the problem. You will write a project proposal on the Project Minerva forum. Open a thread with your Firstname\_Lastname (or the Firstname\_Lastname of every person in your group) as the topic. The teachers will guide you in delineating an appropriate project if necessary.

Describe a problem or program you want to write in Python. The initial post is mainly for brainstorming purposes. We will discuss and give suggestions to delineate the project.

If you have trouble thinking of a project, think about a personal/academic field of interest, identify a task that is repetitive, requires a lot of counting or data processing. Try to make a vague idea as specific as possible.

***Deadline: 20 November 2017, 23:59h***.

### Step 2: Problem analysis

When you have a well-delineated problem, you will move to step 2 of the project proposal which involves breaking down the problem in several steps.

In your forum thread make a post with:

* Introduce the problem you want to solve.
* Problem analysis:
  + Input:
    - Describe the data requirements:
      * Do you need data? Will you use an existing corpus/dataset or will you collect your own?
      * What is the form and encoding of the data?
      * Where will you get the data if you do not have it?
      * How will you collect the data?
    - Describe if your program requires user input.
  + Output:
    - Describe the expected result of the program.
    - Describe which information you write to file.
  + Processing:
    - Describe the expected steps involved in going from input to output.
* Give an example of typical input and the output you expect.

We will discuss the project proposal together on the forum and in the project class sessions. When the project discussion is completed you may write a brief project description based on the schema above to a file (.txt, .pdf, .docx) to be included in the final project folder.

***Deadline: 27 November 2017, 23:59h***.

## B. Project guidelines:

* Variable and function names should be in English. (This will also help when looking online for help.)
* Comments: Write clear and concise comments in the source code that explain what a function does. Try to avoid commenting where functionality is really obvious
* We advise to keep the project text-based, i.e. do not try to include a fancy GUI. GUIs are often completed and require a lot of work. Try to work with text-based inputs and outputs.

## C. Submission

Submit the project folder as zipfile named Firstname\_Lastname\_project.zip file to the relevant Minerva gradebook before the deadline.

Your project zip should contain:

* A file with the project proposal (.txt, .pdf, .docx).
* A subfolder named "src" containing all Python scripts.
  + The "src" folder should also contain a README.txt file with a brief explanation of the project, how to use the code and run the program, and any further comments.
* A subfolder named "data" if you make use of any files.

***Deadline: 24 January 2018, 23:59h***.

## D. Evaluation

The final project will be scored on four criteria:

* The scope of the problem (this does not equal the amount of lines of code!).
* Code quality:
  + Program design: Is the problem solved correctly? Is the program designed in a well-planned and thought out manner?
  + Program execution: Does it run correctly without issue? Does it provide correct output consistently? Does it work on unexpected inputs? Testing with many different inputs is adviced.
* Readability: Is the code well-written and commented. Does it have clear and understable variable names?
* Process: Did you independently use the course material and online sources to get to good solutions?

## E. Example projects:

* Price comparison tool: Compares prices on several webstores for a certain book by webscraping.
* Interactive vocabulary learning tool: Reads a vocabulary list, asks a students for the translation of each word, and scores the student.
* Post-processing of machine translation output: Process KantanMT machine translation output to correct errors in translation of numbers.
* Crossword solver: lists possible words based on length and filled in letters.
* Lexical relation tool: list synonyms and antonyms, together with their gloss for a list of words.
* A roulettes gambling game.

General techniques and topics:

* Web crawling in order to collect a dataset/text corpus.
* Enriching text by linguistic processing and analysis of some sort.
* Automated corpus analysis.