Assignment 2: Coding the Humanities

Coding Assignment

January, Block III, 2017/2018

22 January 2018

1 Assignment Description: Text Mining in Python.

The first three weeks consisted of a broad introduction to Python programming, with a focus on reading, processing and interpreting text data. In this exercise you will be asked to apply all the techniques you learned to a data set of your choice (but preferably a corpus you are familiar with): this can be a collection of novels on Gutenberg, data retrieved via the Guardian API, or just a website you find particularly inviting to analyse in depth. Before you start downloading the data and writing pieces of code, formulate a research question and motivate your choice: what data are you investigating and what questions can be answered by analyzing these texts quantitatively? Only then (after defining the goal of the project) proceed with interrogating your data.

2 Assignment Requirements and Instructions.

The paper should meet the following formal requirements:

- Process a number of files: either download a bunch of documents from the Web or retrieve them via an API.
- Extract relevant statistics from those files; make comparisons, find relevant fragments and patterns in your text: The sky is the limit here. Try to answer your research questions as well as possible with the tools you acquired during this course.
- Present the resulting information to the user: for example by showing frequency tables or plotting time lines.
- Store the information in a useful format (optional).

The paper should be handed in as a **Jupyter Notebook**. Make use of the Markdown cells to write the introduction, the conclusion and to explain/comment on the structure of your program. Use the Code cells for running your actual program.

The paper should have the following structure:

- Introduction: Explain your initial research question, the scope and content of your data (and of course how these relate to the issue you'd like to investigate). Preferably, contextualize the question with relevant literature (ca. 250 words).
- Writing the program: Explain the goal and structure of your program. The section should combine the Markdown with Code cells (ca. 500 words excluding the code).
- Analysis and discussion of the results: Reflect on the results obtained, their value and limitations (ca. 500 words).
- Conclusion (whatever amounts of words you need).

You can form a group of maximum to people.

- multiply word count by 1.75 if your group consists of two participants.
- multiply word count by 2.5 if your group consists of three participants.

3 Assignment Procedure

- Wednesday 24: Form groups and pitch your project to me in class.
- Monday 29 January to Friday 2 February: Time slots for individual feedback will be posted on blackboard.
- **Monday 5 February, 23:59**h: Final Paper for Assignment 2 due. To be uploaded on Blackboard.

4 Grading

4.1 Aspects

This assignment counts for 60% of the total grade. Grading will assess the following aspects:

- The project should be handed in as a Notebook in which you use Markdown and Code cells correctly.
- The code should run (I should be able to rerun the whole program).
- Code is well commented and explained.
- Correct application of concepts: list, dictionaries, etc.
- Originality: relates to both the data and the application.
- **Bonus** if you can make a structured program using well-crafted functions.

4.2 Scheme

The assignment will be graded using the following schema:

- 10 - The paper can be published without any additional work. Perfect!
- 8 - 9: Excellent. Highly original work. You pushed the tools you learned in this course to their limits.
- 6 - 7: You did very well in reproducing existing findings/concepts, you code is reasonably efficient, but the paper lacks a innovation.
- 5: You have written the required amount of words and code, without making fatal errors.
- 2 - 4: You did an attempt, but not very seriously. Your paper either lacks a personal contribution, and/or does not meet the formal requirements.
- 1: You filled out your name and student number correctly.

5 Deadline

Hand in your assignment k.beelen@uva.nl on Sunday **February 5 2018, by 23:59** latest. You can upload it on Blackboard as well.

6 Literature

Useful literature and further reading:

Downey, Allen. Think Python. "O'Reilly Media, Inc.", 2012.
Lutz, Mark. Learning Python: Powerful Object-Oriented Programming. "O'Reilly Media, Inc.", 2013.
Montfort, Nick. Exploratory Programming for the Arts and Humanities. MIT Press, 2016.
Sweigart, Al. Automate the Boring Stuff with Python: Practical Programming for Total Beginners. No Starch Press, 2015.