**Hacking the Humanities: Programming and Analysis**
**Mondays 13:15-15:00 Lipsius 126/127**
**Fall 2018**

Instructor: Paul Vierthaler
Office: PJ Veth (Herbarium) T0.14
Office Hours: Tuesdays 13:00-16:00 and by appointment
p.a.vierthaler@hum.leidenuniv.nl

Teaching Assistant: Martin Kroon (T0.10)
m.s.kroon@hum.leidenuniv.nl

This course provides students with an introduction to the digital humanities. It builds a firm foundation of technical skills and introduces students to the basic questions in the field. Although we will engage with theoretical work on the digital humanities and read secondary materials, this is primarily a methods course. While a basic familiarity with computers is recommended, students will not need to have any prior programming or command line experience.

All of the tools used in this course are open-source or freely available on the internet. We will be using Anaconda (a scientific computing distribution of the Python programming language), Processing, and other open source software.

In this course, students will learn to identify and acquire useful materials, program ad-hoc tools, and use digital analytical techniques where appropriate. The course will culminate in a 2,000 to 3,000 word research paper and a short ten-minute presentation. This paper must contain at least two figures. Alternatively, students can develop an online project (with a similar amount of content) in consultation with me. Students should meet with me by October 31st to discuss ideas for final papers/projects.

Course materials will be distributed through GitHub and Blackboard.

Over the course of the semester, I will provide students with assignments that will help guide you through the important concepts. Each assignment will require the student to write a short program that implements techniques from previous sessions and an essay describing the program and its purpose. Students should complete the assignment on her/his/their own. You are encouraged to search the internet for help in finishing the assignments. If you find a solution to the problem somewhere online, you can use it but post the source and explain in detail why the solution works.

Class attendance is required. The information we cover is cumulative, so if you do not attend, it will be very easy to fall behind. Attendance and active participation are obligatory for seminars. Students are required to prepare for and attend all sessions. The convenors need to be informed without delay of any classes missed for a good reason (i.e. due to unforeseen circumstances such as illness, family issues, problems with residence permits, the Dutch railways in winter, etc.). In these cases it is up to the discretion of the convener(s) of the course whether or not the missed class will have to be made up with an extra assignment. The maximum of such absences during a

semester is two. Being absent without notification and/or more than two times can result in a failing grade for the course.

Each session will include a lecture, a fifteen-minute break, and a practicum.

Prior to each session please do the readings and watch the applicable tutorial videos.

Video Tutorials are available on YouTube.

**Computers in the lab have the necessary software installed, but it may be easier for you to use your own laptop. If you have one, please bring it to class.**

**Grading Rubric:**
Attendance and Participation: 10 percent
Assignments: 30 percent
Final Project (paper or online project): 50 percent
Project presentation: 10 percent

Final mark for the course is established by determining the weighted average.

**Late Assignments/Late Papers:** will be penalized by a 10% reduction for each 24-hour period it is late. After one calendar week, the assignment will not be accepted.

**Writing Expectations:** While content is the most important component of your work, a portion of your grade on each assignment will be for style and understandability. There is a writing centre at Leiden. Feel free to avail yourself of it!

**Resit:** Students who have scored an overall insufficient grade for the course may take a resit for the assignments and the Paper (or online project):

comprehensive take-home exam: (50 percent of grade)
Paper (or online project): (50 percent of grade)

**A Word About Plagiarism**

You must document all of your source material. If you take any text from somebody else, you must make it clear the text is being quoted and where the text comes from. You must also cite any sources from which you obtain numbers, ideas, or other material. If you have any questions about what does or does not constitute plagiarism, ask! Plagiarism is a serious offense and will not be treated lightly. Fortunately, it is also easy to avoid and if you are the least bit careful about giving credit where credit is due you should not run into any problems (thanks to Alfred E Guy, Jr. for this statement).

**Course Schedule**

September 10: Course Introduction
What are the Digital Humanities and why build your own tools?
Programming Languages for DH: Python, Ruby, R, Processing
Pre-built tools: Gephi, MALLET, qGIS

Installing and running the programs:
Anaconda, Processing
Introducing the command line

**Part 1: Intro to Programming for the Digital Humanities and Text Analysis**

September 17: Strings, Integers, Floats, and Math
Building and running basic programs

Reading:
Anne Burdick,et.al, "From Humanities to Digital Humanities"
https://mitpress.mit.edu/books/digitalhumanities pages 1-26
You can download the open access version for free from the left sidebar

September 24: Computational Approaches to Literature
Lists, Booleans, Loops
Basic analysis of a small corpus

Reading:
Stephen Ramsay, *An Algorithmic Criticism*: 1-17
Matthew Jockers, *Macroanalysis*: Part 1: Foundation, 3-32

October 1: Dictionaries, Libraries, NLTK, Persisting Data

Reading:
Susan Hockey, "The History of Humanities Computing" in *A Companion to Digital Humanities*
http://www.digitalhumanities.org/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-2-1

"Text Analysis with NLTK Cheatsheet:"
Posted on Blackboard

October 8: Regular Expressions (lecture Martin Kroon)
Reading:
"Understanding Regular Expressions" *The Programming Historian:*

https://programminghistorian.org/en/lessons/understanding-regular-expressions

*Dive into Python*: "Chapter 7: Regular Expressions."
http://www.diveintopython.net/regular_expressions/

October 15: Git, GitHub, Collaborative coding, and Fixing Errors

Reading:

Will be distributed via Blackboard

**Part 2: Data Analysis and Visualization**

October 29: Data Analysis with Pandas

Finding structured data, analyzing structured data with Pandas
Import and analyze a structured dataset.

Reading:

Christof Schöch, "Big? Smart? Clean? Messy? Data in the Humanities," *Journal of Digital Humanities* Vol.2, No.3, Summer 2013
http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/

"10 minutes to Pandas:"
http://pandas.pydata.org/pandas-docs/dev/10min.html#min

November 5: Visualization in Python

Why visualize data (and how do you do it in Python)?

Reading:

Nathan Yau, *Visualize This:* Chapter 1
Joanna Drucker, "Humanities Approaches to Graphical Display"
http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html

**Part 3: More Advanced Text Analysis**

November 12: Stylometry and Python

Reading:

Christof Schoch, "Principal Component Analysis for Literary Stylistics:"
http://dragonfly.hypotheses.org/472

Peter Turney and Patrick Pantel, "From Frequency to Meaning: Vector Space Models of Semantics," *Journal of Artificial Intelligence Research* 37 (2010): 141-

November 19: Topic Modelling (Lecture Martin Kroon)
Topic modeling and its uses in the humanities
Running a topic model with MALLET and gensim

Reading:
Ted Underwood, "Topic-modelling made just simple enough:"
http://tedunderwood.com/2012/04/07/topic-modeling-made-just-simple-enough/

**Part 4: Networks and Maps**

November 26: Visualization in Processing (making Animations)
Reading:
Read through the processing overview: https://processing.org/overview/

Do tutorial at http://hello.processing.org/ (this will take approximately an hour)

December 3: APIs and Webscraping:
When, where, why (and how) do we ethically scrape the web?

Reading:
Wendy Hsu, "Digital Ethnography Toward Augmented Empiricism: A New Methodological Framework," *Journal of Digital Humanities*, Vol.3, No. 1, Spring 2014
http://journalofdigitalhumanities.org/3-1/digital-ethnography-toward-augmented-empiricism-by-wendy-hsu/

Look over the BeautifulSoup documentation:
http://www.crummy.com/software/BeautifulSoup/bs4/doc/

**Wrap up**

December 10: Project presentations

Final paper due January 5th.