

The ABC of Computational Text Analysis

#3 SETTING UP YOUR DEVELOPMENT ENVIRONMENT

Alex Flückiger

Faculty of Humanities and Social Sciences
University of Lucerne

11 March 2022

Faculty of Humanities and Social Sciences
University of Lucerne

Outline

- research workflow 
- interacting with computers 
- installation of programs 

Recap last Lecture

- textual data is challenging
- machine-readable data

A close-up photograph of a large pile of jigsaw puzzle pieces. The pieces are of various colors, including shades of blue, yellow, orange, red, green, and brown. They are scattered across the frame, creating a textured, organic pattern. The lighting is bright, highlighting the edges and surfaces of the individual puzzle pieces.

The Zen of Organizing

Research means Organizing

Computational approach

- scripts as documentation 

- efficient automation 

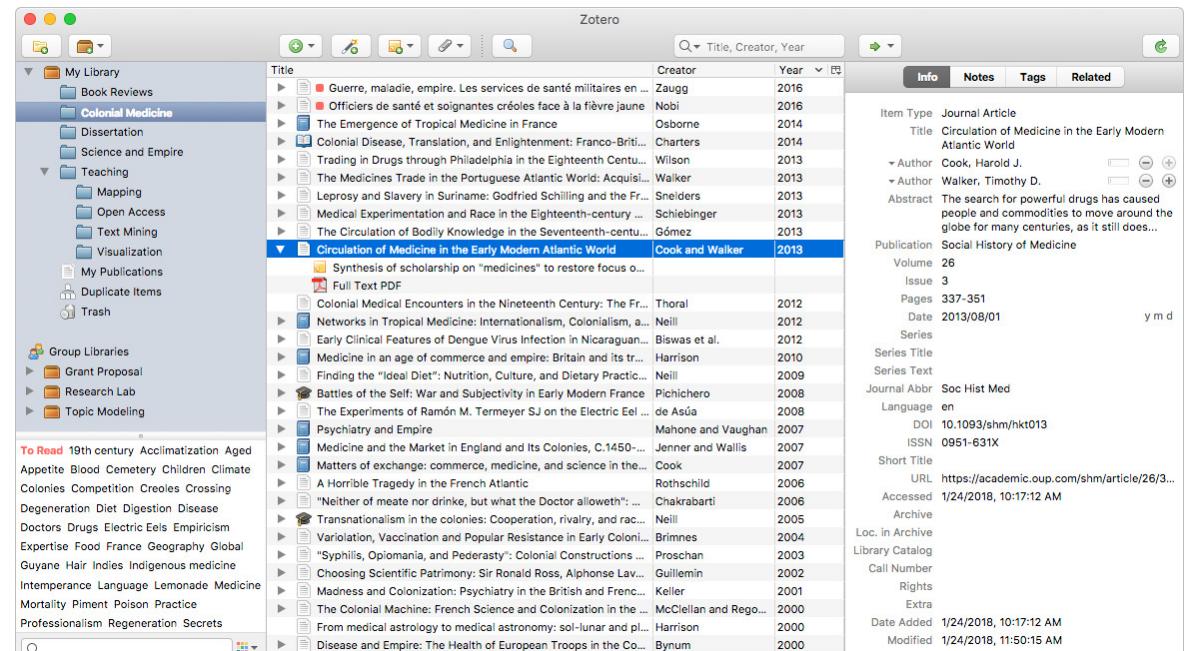
"don't repeat yourself"

- less error-prone 

- reproducible 

Organizing Literature

- manage literature in one place
- generating bibliographies
- collect with a click



Zotero

Computers

Two Trends in Computing

- everyday, mobile, cloud-centered, touch/click-based model
 - single-purpose application*
 - smartphones*
- engineering, open-source, machine-oriented model
 - powerful due to modularity*
 - workflow for data science*



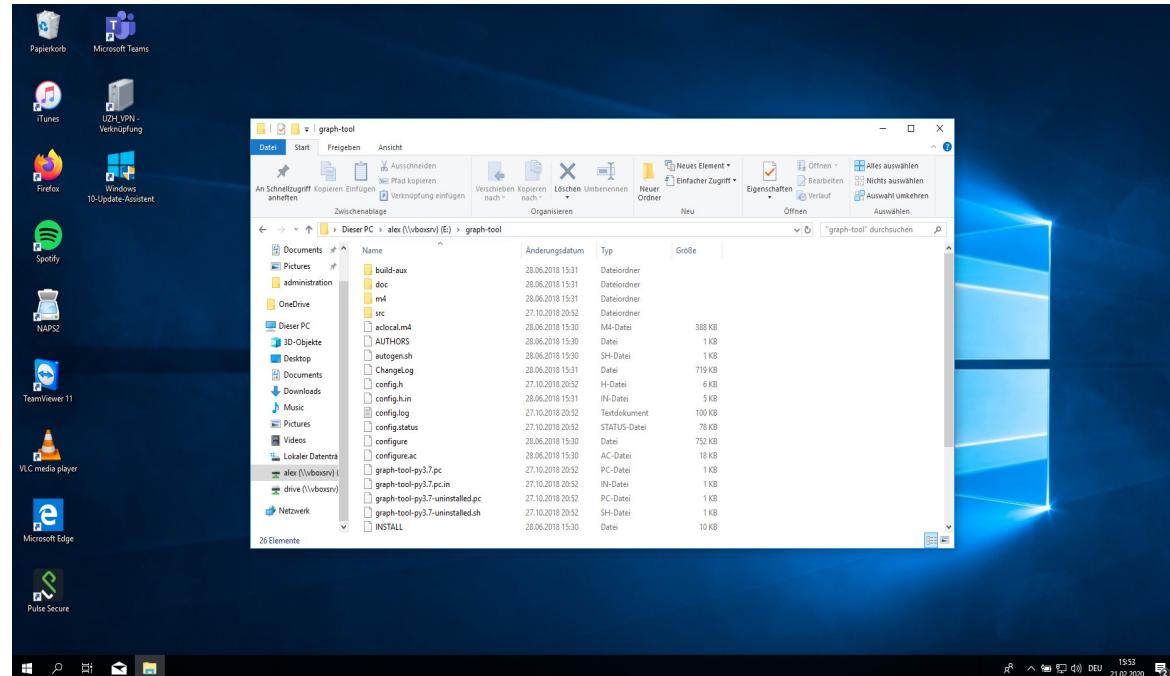
(see also Healy 2019)

Operating Systems (OS)

- Microsoft Windows
- macOS
- Linux / Unix 😎

User Interfaces

graphical user interface (GUI)



command line interface (CLI)

```
[alex@alex-pc graph-tool]$ ls -lh
total 2.6M
-rw-r--r-- 1 alex alex 388K 28. Jun 2018 aclocal.m4
-rw-r--r-- 1 alex alex 73 28. Jun 2018 AUTHORS
-rwxr-xr-x 1 alex alex 963 28. Jun 2018 autogen.sh
drwxr-xr-x 2 alex alex 4.0K 28. Jun 2018 build-aux
-rw-r--r-- 1 alex alex 719K 28. Jun 2018 ChangeLog
-rw-r--r-- 1 alex alex 5.4K 27. Okt 2018 config.h
-rw-r--r-- 1 alex alex 4.4K 28. Jun 2018 config.h.in
-rw-r--r-- 1 alex alex 100K 27. Okt 2018 config.log
-rwxr-xr-x 1 alex alex 78K 27. Okt 2018 config.status
-rwxr-xr-x 1 alex alex 752K 28. Jun 2018 configure
-rw-r--r-- 1 alex alex 18K 28. Jun 2018 configure.ac
drwxr-xr-x 6 alex alex 4.0K 28. Jun 2018 doc
-rw-r--r-- 1 alex alex 692 27. Okt 2018 graph-tool.py3.7.pc
-rw-r--r-- 1 alex alex 451 27. Okt 2018 graph-tool.py3.7.pc.in
-rw-r--r-- 1 alex alex 711 27. Okt 2018 graph-tool.py3.7-uninstalled.pc
-rw-r--r-- 1 alex alex 1017 27. Okt 2018 graph-tool.py3.7-uninstalled.sh
-rw-r--r-- 1 alex alex 9.3K 28. Jun 2018 INSTALL
-rwxr-xr-x 1 alex alex 344K 27. Okt 2018 libtool
-rw-r--r-- 1 alex alex 35K 28. Jun 2018 LICENSE
drwxr-xr-x 2 alex alex 4.0K 28. Jun 2018 m4
-rw-r--r-- 1 alex alex 34K 27. Okt 2018 Makefile
-rw-r--r-- 1 alex alex 1.5K 28. Jun 2018 Makefile.am
-rw-r--r-- 1 alex alex 33K 28. Jun 2018 Makefile.in
-rw-r--r-- 1 alex alex 2.5K 28. Jun 2018 README.md
drwxr-xr-x 5 alex alex 4.0K 27. Okt 2018 src
-rw-r--r-- 1 alex alex 23 27. Okt 2018 stamp-h1
[alex@alex-pc graph-tool]$
```

“An awesome Programmer saves the World”

B

Human-Machine Interaction

- text commands instead of clicks
- similar to a human languages
 - syntax (form)*
 - semantics (meaning)*
- computers are literalists

Programming

- command languages
Bourne-Again shell (Bash), since 1989
- programming languages
Python, R, Java, C++ etc.
- instructions → program ≡ algorithm
cooking

Package Manager

- automation of software installation
 - system: apt, Homebrew*
 - Python: conda, pip*
- software dependencies
- similar to app stores

Open-Source is a Mindset

Standing on the shoulders of giants.

- free + open software
- cross-platform portability
- collaboration
 - share + reuse*
- helpful community

Resources

- Stack Overflow
where you get answers to technical questions
- GitHub
where you find open software

Learning by doing, doing by googleing.

Set up your System

Backup

You don't need it until you desperately need it!

Setting up your Development Environment

Installation

- Anaconda Platform with Python 3.8
packages for NLP + visualization
- Tesseract
- various Bash tools

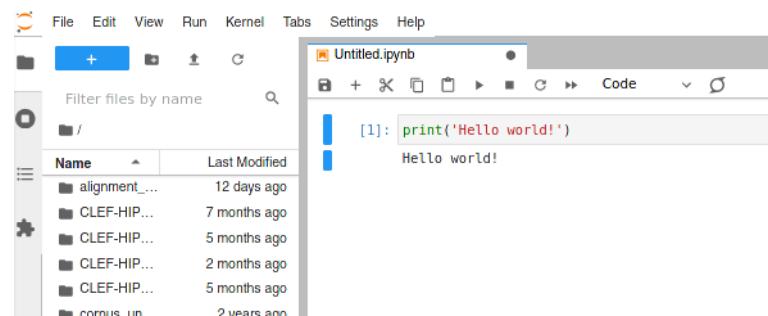
 Follow the [installation guide](#) for your OS.



Relax.
It takes a while.

JupyterLab

- a web-based development environment
- interactive analysis
- cross-platform



First steps in Python

Learn Python with interactive tutorials

- [Python Principles](#)
- [LearnPython](#)

Readings

optional: pimp your workflow

Healy, Kieran. 2019. "The Plain Person's Guide to Plain Text Social Science." [online](#).



Questions?

References

Healy, Kieran. 2019. "The Plain Person's Guide to Plain Text Social Science."
<https://kieranhealy.org/publications/plain-person-text/>.