

The ABC of Computational Text Analysis




#3 SETTING UP YOUR DEVELOPMENT ENVIRONMENT

Alex Flückiger

Faculty of Humanities and Social Sciences
University of Lucerne

17 March 2022

Outline

- research workflow 
- interacting with computers 
- installation of programs 

Recap last Lecture





- textual data is challenging
- machine-readable data



The Zen of Organizing

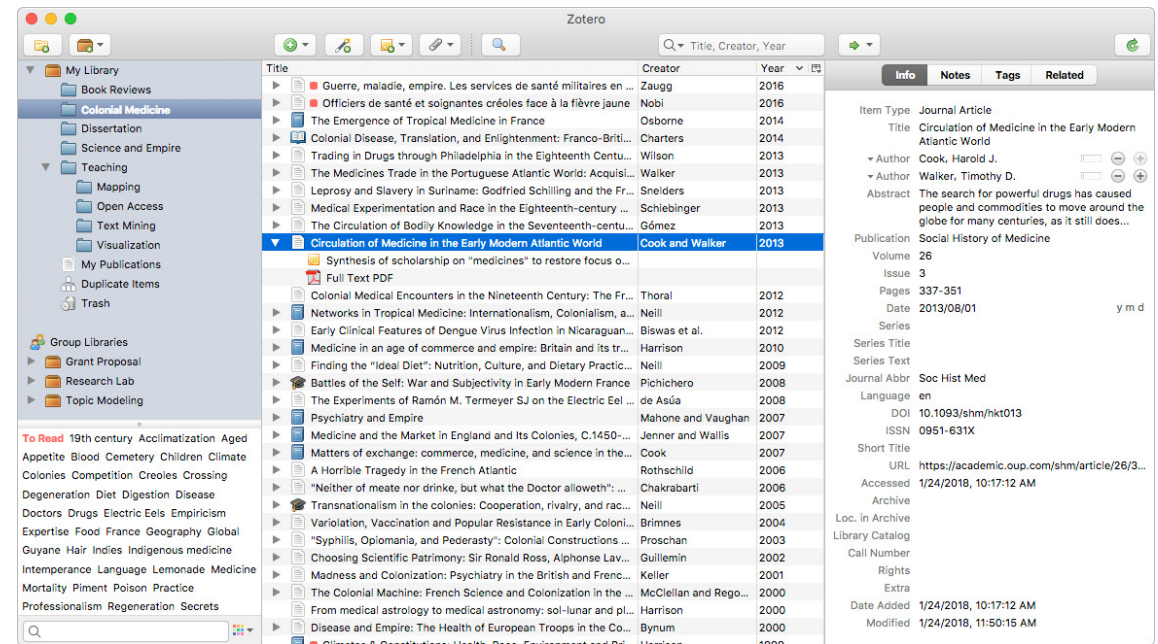
Research means Organizing

How a computational approach helps

- scripts as documentation 
- efficient automation 
“don't repeat yourself”
- less error-prone  
- reproducible 100

Organizing Literature

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



Zotero is a free, open-source software for managing scientific literature

Computers

Two Trends in Computing

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

(see also Healy 2019)



Apple Original iPhone (2007)

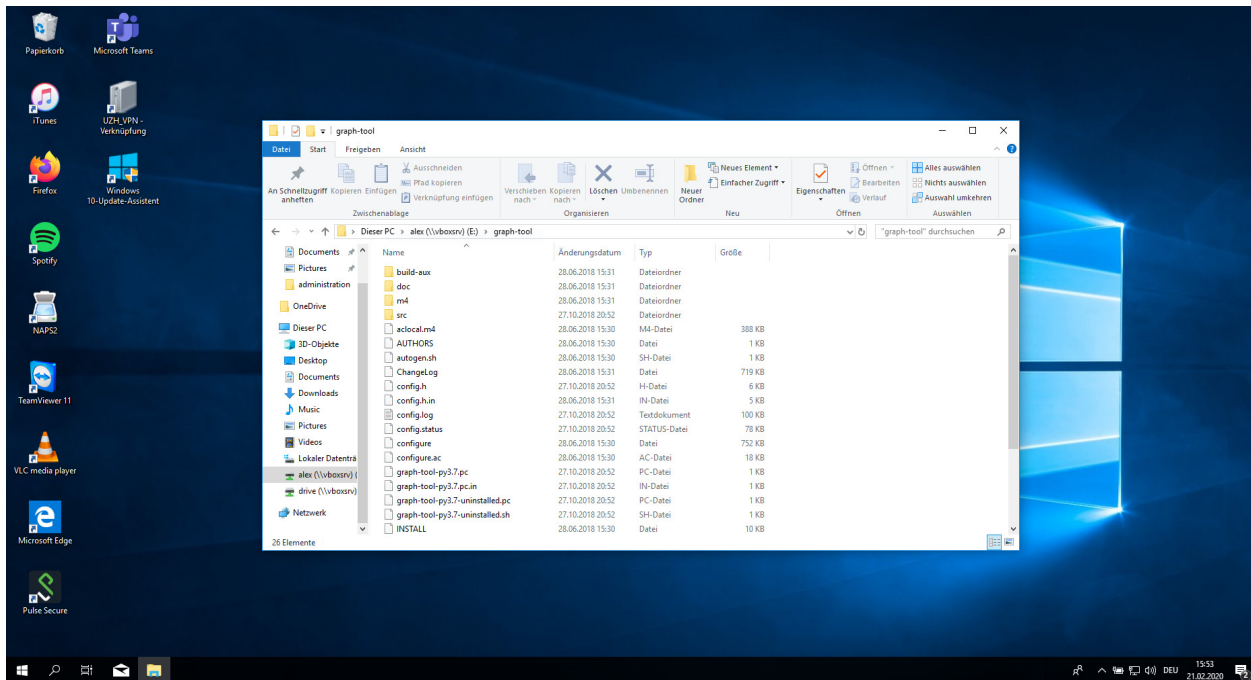
Operating Systems (OS)

Between hardware and programs

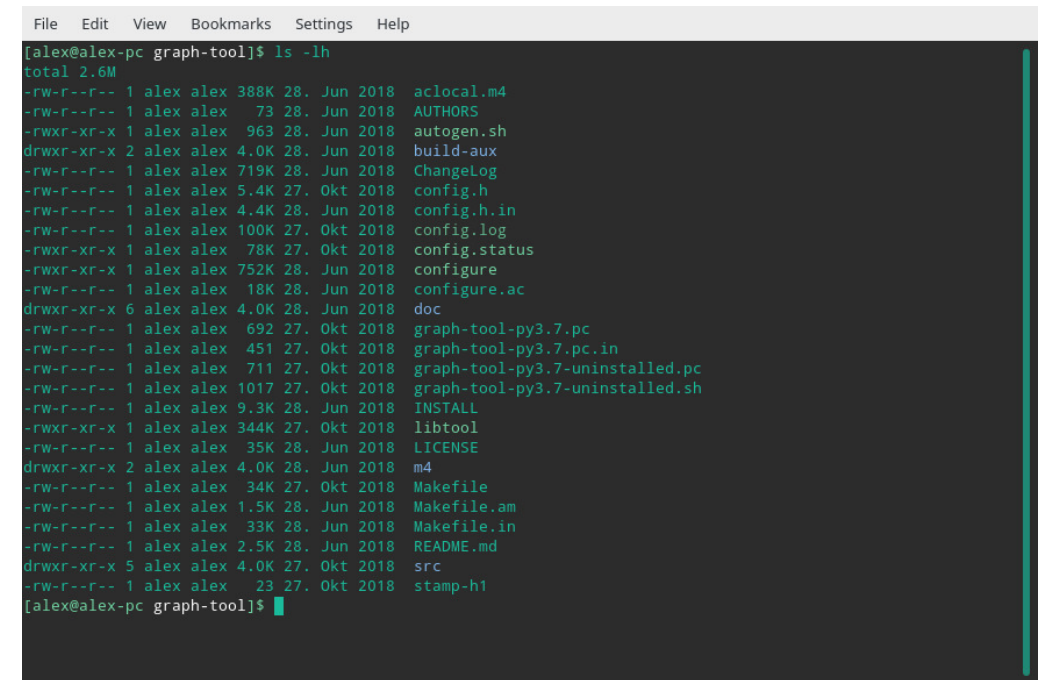
- Microsoft Windows
- macOS
- Linux-based systems 🧐

User Interfaces

graphical user interface (GUI)



command line interface (CLI)



“An awesome Programmer saves the World”
8

Human-Machine Interaction

How to tell the computer efficiently

- text commands instead of clicks
- similar to 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** \equiv **algorithm**
cooking

Package Manager

- automation of software installation
 - system: apt, Homebrew
 - Python: pip, conda
- 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 everyone is using

- **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

- Python 3.8
 - packages for NLP + visualization
- Tesseract
- various Bash tools
- VS Code Editor

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

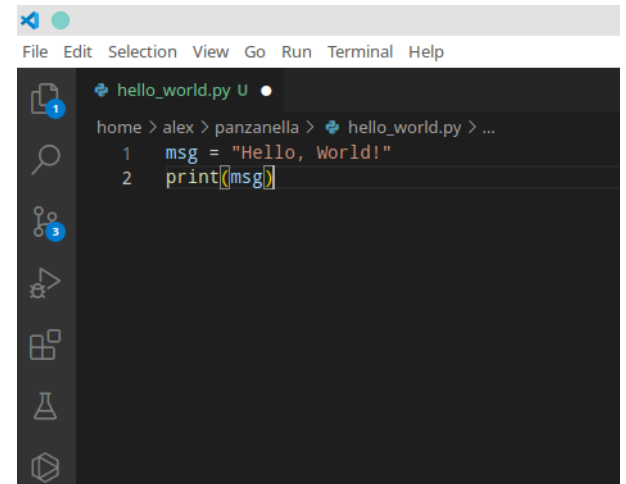


Relax.
It takes a while.

VS Code Editor

The Microsoft Word for coding

- powerful integrated development environment (IDE)
- cross-platform
- interactive analysis

A screenshot of the Visual Studio Code (VS Code) editor interface. The window title is "hello_world.py U". The menu bar includes "File", "Edit", "Selection", "View", "Go", "Run", "Terminal", and "Help". The editor shows a Python script with two lines of code: "1 msg = 'Hello, World!'" and "2 print(msg)". The terminal output shows the command "home > alex > panzanella > hello_world.py > ..." and the output "Hello, World!". The left sidebar contains icons for Explorer, Search, Source Control, Run and Debug, Extensions, Testing, and Docker.

Write your first Python script in [VS Code](#)

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/>.