

The ABC of Computational Text Analysis

#3 SETTING UP YOUR DEVELOPMENT ENVIRONMENT

Alex Flückiger

Faculty of Humanities and Social Sciences

University of Lucerne

March 17, 2023

Outline

- research workflow 
- interacting with computers 
- installation of programs 

Recap last Lecture

- textual data is challenging
- machine-readable data

A large pile of colorful puzzle pieces, mostly grey, yellow, and orange, scattered across the entire background.

The Zen of Organizing

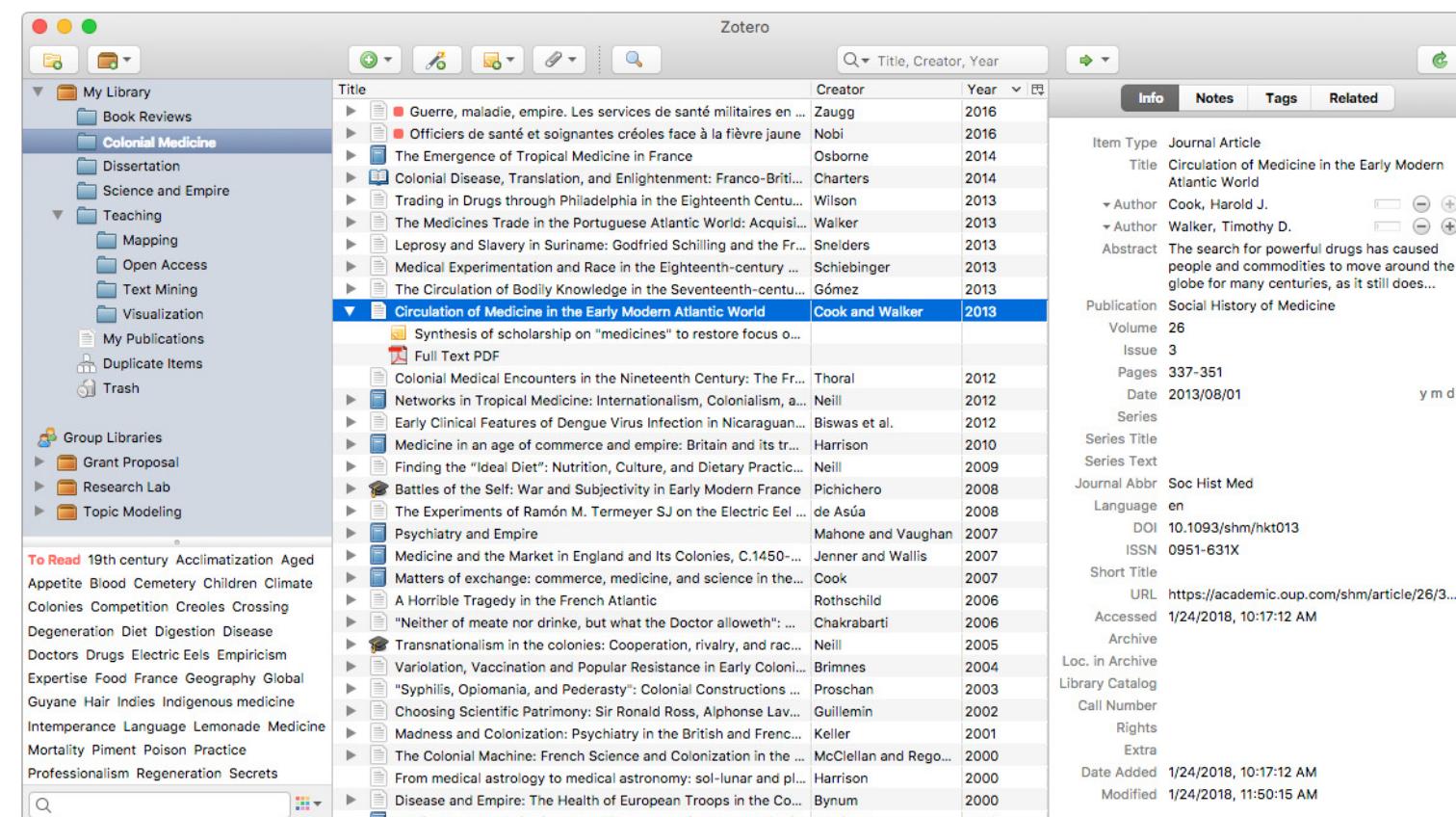
Research means Organizing

How a computational approach helps

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

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)

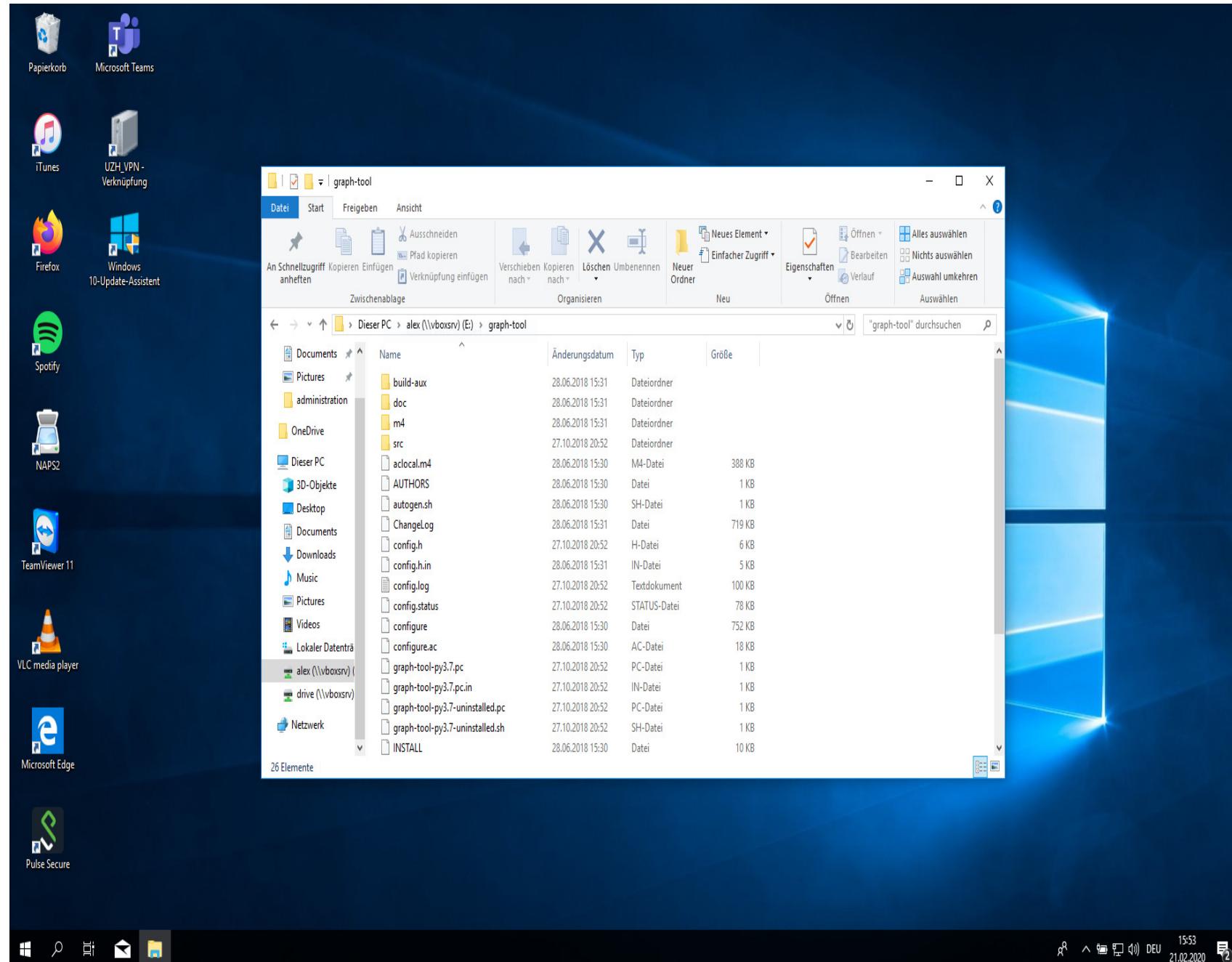
Operating Systems (OS)

Between hardware and programs

- Microsoft Windows
- macOS
- Linux-based systems 😎

User Interfaces

graphical user interface (GUI)



command line interface (CLI)

A screenshot of a terminal window with a dark theme. The window title is '[alex@alex-laptop KED2023]\$. ls -lht KED2023/'. The terminal displays a list of files and directories in the 'KED2023' directory. The output is as follows:

```
[alex@alex-laptop KED2023]$ ls -lht KED2023/
total 800K
-rw-r--r-- 1 alex alex 1.3K Feb 3 16:22 _quarto.yml
-rw-r--r-- 1 alex alex 734K Feb 3 16:06 KED2023_syllabus.pdf
-rw-r--r-- 1 alex alex 2.1K Feb 3 15:42 schedule.qmd
drwxr-xr-x 7 alex alex 4.0K Feb 3 15:13 materials
-rw-r--r-- 1 alex alex 15K Jan 27 18:54 lectures.qmd
drwxr-xr-x 8 alex alex 4.0K Jan 27 16:52 lectures
-rw-r--r-- 1 alex alex 5.6K Jan 17 13:34 assignments.qmd
-rw-r--r-- 1 alex alex 2.0K Jan 17 13:34 index.qmd
-rw-r--r-- 1 alex alex 2.4K Jan 17 13:34 materials.qmd
-rw-r--r-- 1 alex alex 181 Jan 16 16:44 dark_styles.scss
-rw-r--r-- 1 alex alex 120 Jan 16 16:33 light_styles.scss
drwxr-xr-x 3 alex alex 4.0K Jan 16 15:38 _extensions
drwxr-xr-x 5 alex alex 4.0K Oct 21 15:57 assignments
drwxr-xr-x 3 alex alex 4.0K Oct 21 15:57 _images
-rw-r--r-- 1 alex alex 419 Oct 21 15:57 404.html
[alex@alex-laptop KED2023]$
```

The terminal window also shows three tabs at the bottom: 'KED2023: quarto', 'KED2023: zsh', and 'KED2023: bash'. The 'KED2023: bash' tab is currently active.

1

“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

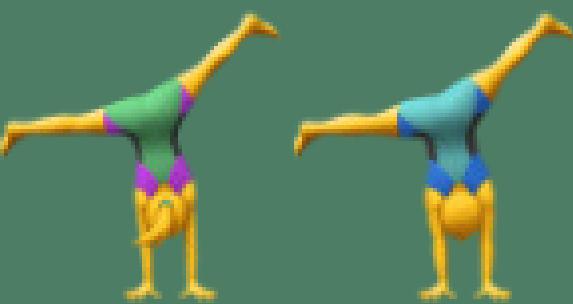
- **Stack Overflow**

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

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

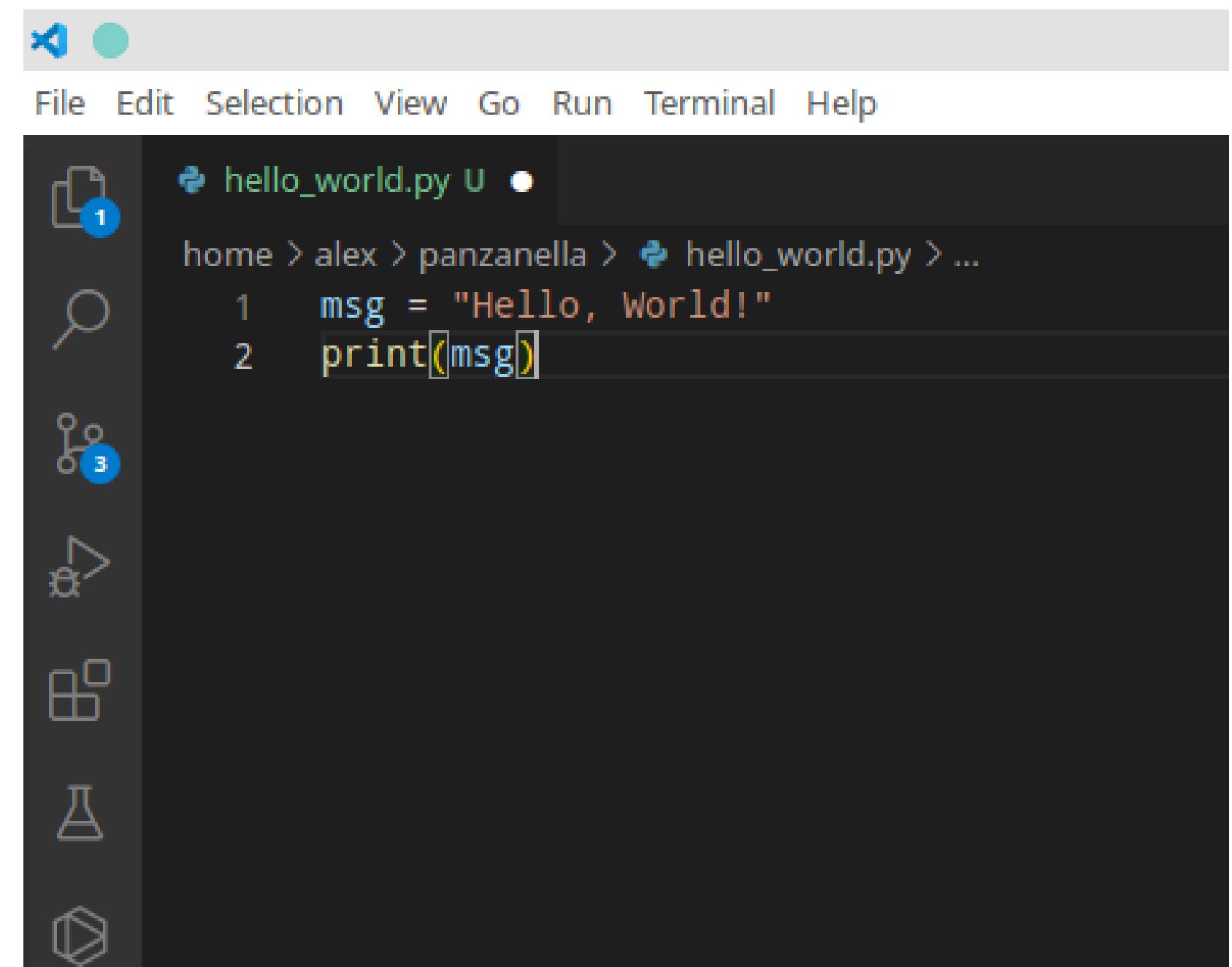


Relax.
It takes a while.

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) interface. The window title is "VS Code". The menu bar includes File, Edit, Selection, View, Go, Run, Terminal, and Help. The left sidebar has icons for file explorer (with 1), search (with 0), symbols (with 3), and other developer tools. The main editor area shows a Python script named "hello_world.py":

```
home > alex > panzanella > hello_world.py > ...
1 msg = "Hello, World!"
2 print(msg)
```

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/>.