

The ABC of Computational Text Analysis

#3 SETTING UP YOUR DEVELOPMENT ENVIRONMENT

Alex Flückiger

Faculty of Humanities and Social Sciences
University of Lucerne

09 March 2023

Outline

- research workflow 
- interacting with computers 
- installation of programs 

Recap last Lecture

- textual data is challenging
- machine-readable data

The background of the image is a dense, overlapping pile of numerous puzzle pieces. The pieces are of various colors, including shades of blue, yellow, orange, red, green, and grey. Some pieces have abstract patterns or textures, while others appear to have more recognizable images like foliage or landscapes. The lighting creates highlights and shadows on the edges of the pieces, giving them a three-dimensional appearance.

The Zen of Organizing

Research means Organizing

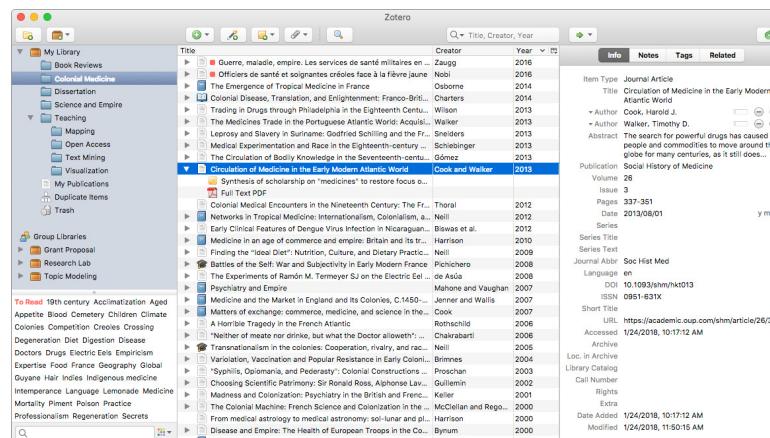
How a computational approach helps

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

Organizing Literature ...

... made easy

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



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)

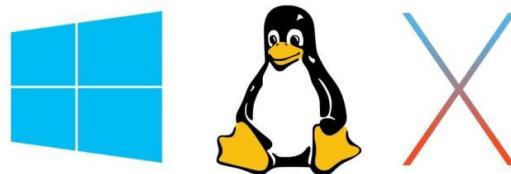


Apple Original iPhone (2007)

Operating Systems (OS)

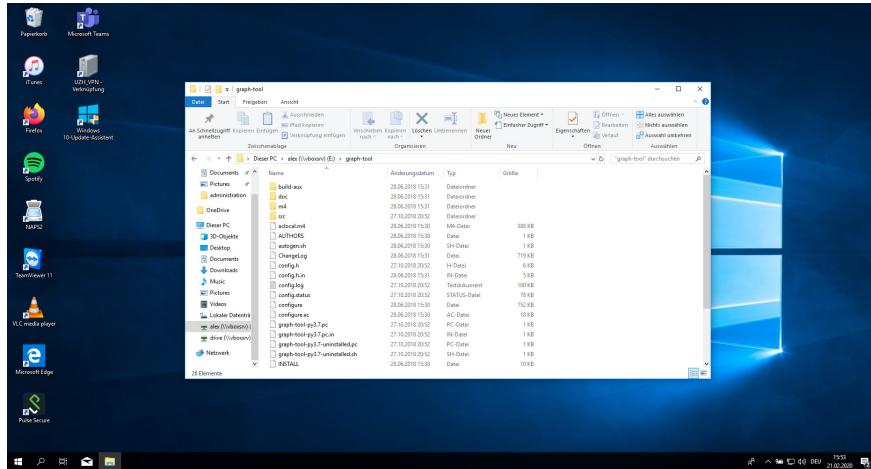
Between hardware and programs

- Microsoft Windows
- macOS
- Linux-based systems 😎



User Interfaces

Graphical User Interface (GUI)



Command Line Interface (CLI)

```
[alex@alex-laptop ~]$ cd KED2023
[alex@alex-laptop KED2023]$ ls -lh
total 340K
drwxr-xr-x 3 alex alex 4.0K Nov  4 18:46 bulk
drwxr-xr-x 7 alex alex 4.0K Mar 16 10:27 docs
-rw-r--r-- 1 alex alex 6.1K Feb 17 15:21 dodo.py
drwxr-xr-x 8 alex alex 4.0K Mar 16 10:28 KED2023
-rw-r--r-- 1 alex alex 1.3K Feb 23 14:05 lecture_intro.md
drwxr-xr-x 2 alex alex 4.0K Feb 17 15:21 lib
-rw-r--r-- 1 alex alex 6.9K Oct 21 15:57 LICENSE
-rw-r--r-- 1 alex alex 288K Feb 17 17:00 poetry.lock
drwxr-xr-x 2 alex alex 4.0K Feb 17 16:05 __pycache__
-rw-r--r-- 1 alex alex 922 Feb 17 17:02 pyproject.toml
-rw-r--r-- 1 alex alex 2.0K Feb 17 15:21 README.md
-rw-r--r-- 1 alex alex 345 Feb 17 15:21 requirements.txt
[alex@alex-laptop KED2023]$
```

1

"An awesome Programmer saves the World"

8



Human-Machine Interaction

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, poetry, 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 3.10
 - packages for NLP + visualization
- Tesseract
- various Bash tools
- VS Code Editor

👉 Follow the [installation guide](#) for your OS + general Python chapter.

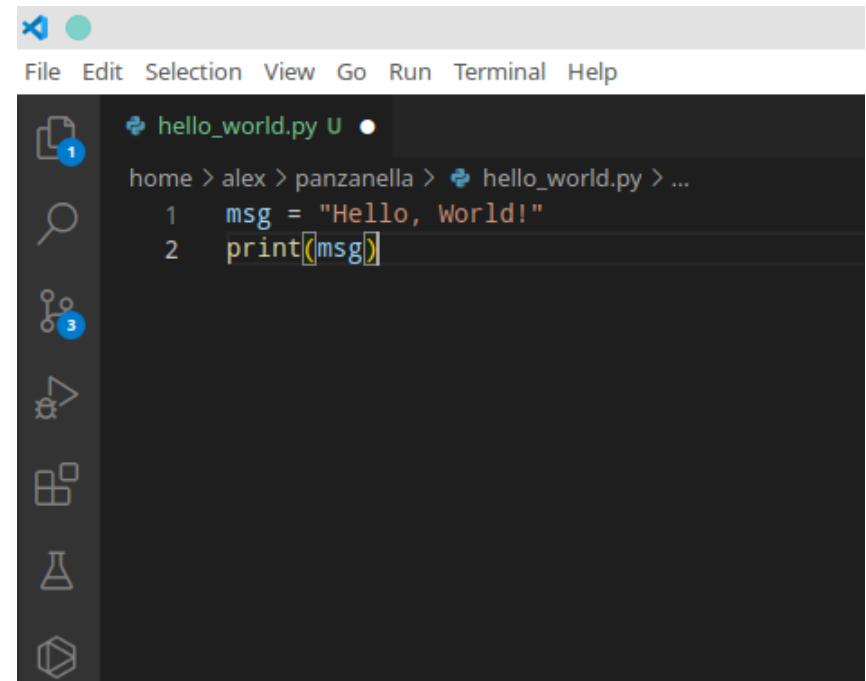
A close-up photograph of a coffee machine dispensing coffee into a clear glass cup. The coffee is being poured from a silver-colored portafilter into the cup, which is positioned on a metal drip tray. The background is blurred, showing the metallic components of the coffee machine.

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 1), and other tools (with 3). The main editor area shows a Python file named "hello_world.py" with the following code:

```
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/>.