

The ABC of Computational Text Analysis

#3 SETTING UP YOUR DEVELOPMENT ENVIRONMENT

Alex Flückiger

Faculty of Humanities and Social Sciences
University of Lucerne

07 March 2024

Outline

- reading + recap
- interacting with computers ✨
- installation of programs 🛠️

Recap last lecture

- textual data is challenging
- machine-readable data
- reproducible research applying computational approach

Reading

Computational Social Science (Lazer et al. 2009)

- data-driven
- network analysis + text analysis
- historical perspective vs. real-time dynamics
- issues: limited access to data and new methods

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)

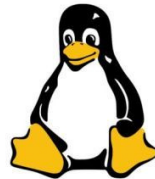
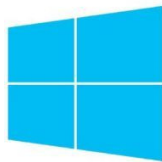


*Apple Original
iPhone (2007)*

Operating Systems (OS)

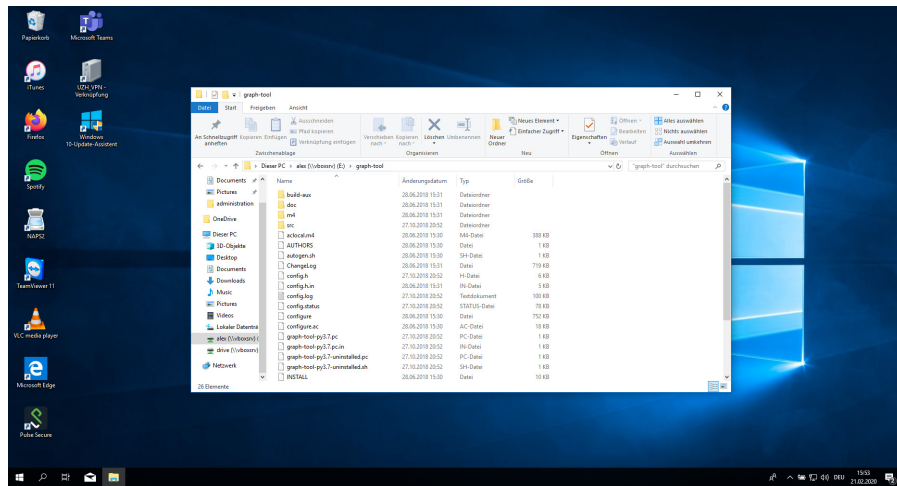
Between hardware and programs

- Microsoft Windows
- macOS
- Linux-based systems 🧐



User Interfaces

Graphical User Interface (GUI)



Command Line Interface (CLI)

```
File Edit View Bookmarks Plugins Settings Help
New Tab Split View Copy Paste Find
```

```
[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]$
```


“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



Organizing Data and Code

- **Git** tracks file changes and allows for version management
- **GitHub** is a popular hosting platform based on Git
 - share code and collaborate
 - repository = project folder



Publishing code and data are key to open science.

Resources everyone is using

- [Stack Overflow](#)

where you get answers to technical questions

- [GitHub](#)

where you find open software

- [ChatGPT](#)

ask anything if you get stuck

Learning by doing, doing by ~~Googleing~~
ChatGPTing.



Set up your System

Backup

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

Setting up your Development Environment

Installation

- Python 3
 - 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. Two streams of dark brown coffee are pouring from the machine's spouts into a clear glass cup. The coffee is creating a frothy head in the cup. The background is a blurred stainless steel surface of the machine.

Relax.
It takes a while.

VS Code Editor

The Microsoft Word for Coding

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



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/>.

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, et al. 2009. "Computational Social Science." *Science* 323 (5915): 721–23.

<https://doi.org/10.1126/science.1167742>.