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 2025

Outline

- recap + discuss paper 📄
- interacting with computers 🪄
- installation of programs 🔨

Recap last lecture

- textual data is challenging 🤯

- machine-readable data 🚧

- reproducible research using 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, smart assistants, augmented reality

- engineering, open-source, machine-oriented approach

    powerful due to modularity
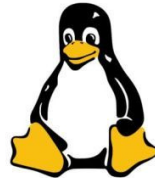
    workflow for data science

(see also Healy 2019)
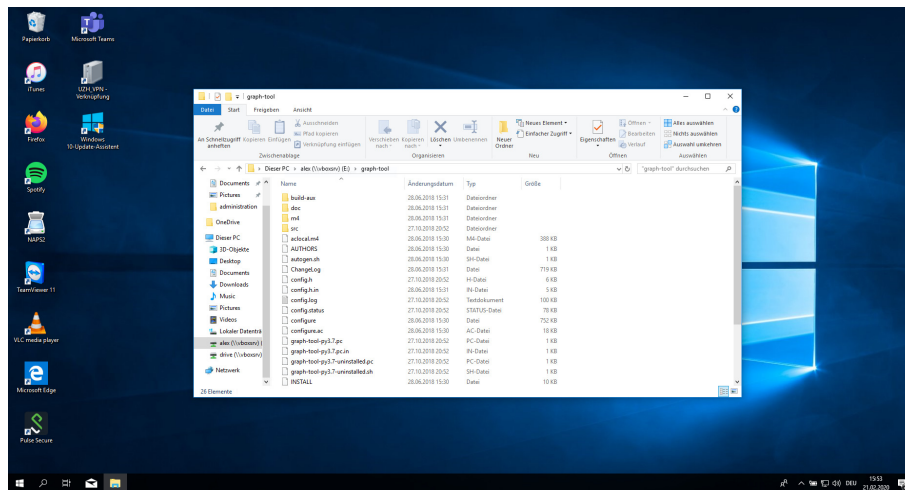

Apple Original iPhone (2007)

Operating Systems (OS)

Between hardware and applications

- Microsoft Windows
- macOS
- Linux-based systems 🤓

User Interfaces (UI)

Graphical User Interface (GUI)

Command Line Interface (CLI)

"An awesome programmer saves the world"

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 ≡ algorithm

  analogous to cooking

Package Manager 📦

App stores for engineers

- automation of software installation

  system: apt, Homebrew

  Python: pip, poetry, conda

- packages have dependencies

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 supports 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 <u>installation guide</u> for your OS + the Python chapter.

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.