The ABC of computational Text Analysis

03: Setting up your Development Environment

Alex Flückiger
12 March 2020

Outline

- research workflow

- interacting with computers

- installation of programs

Recap last Lecture

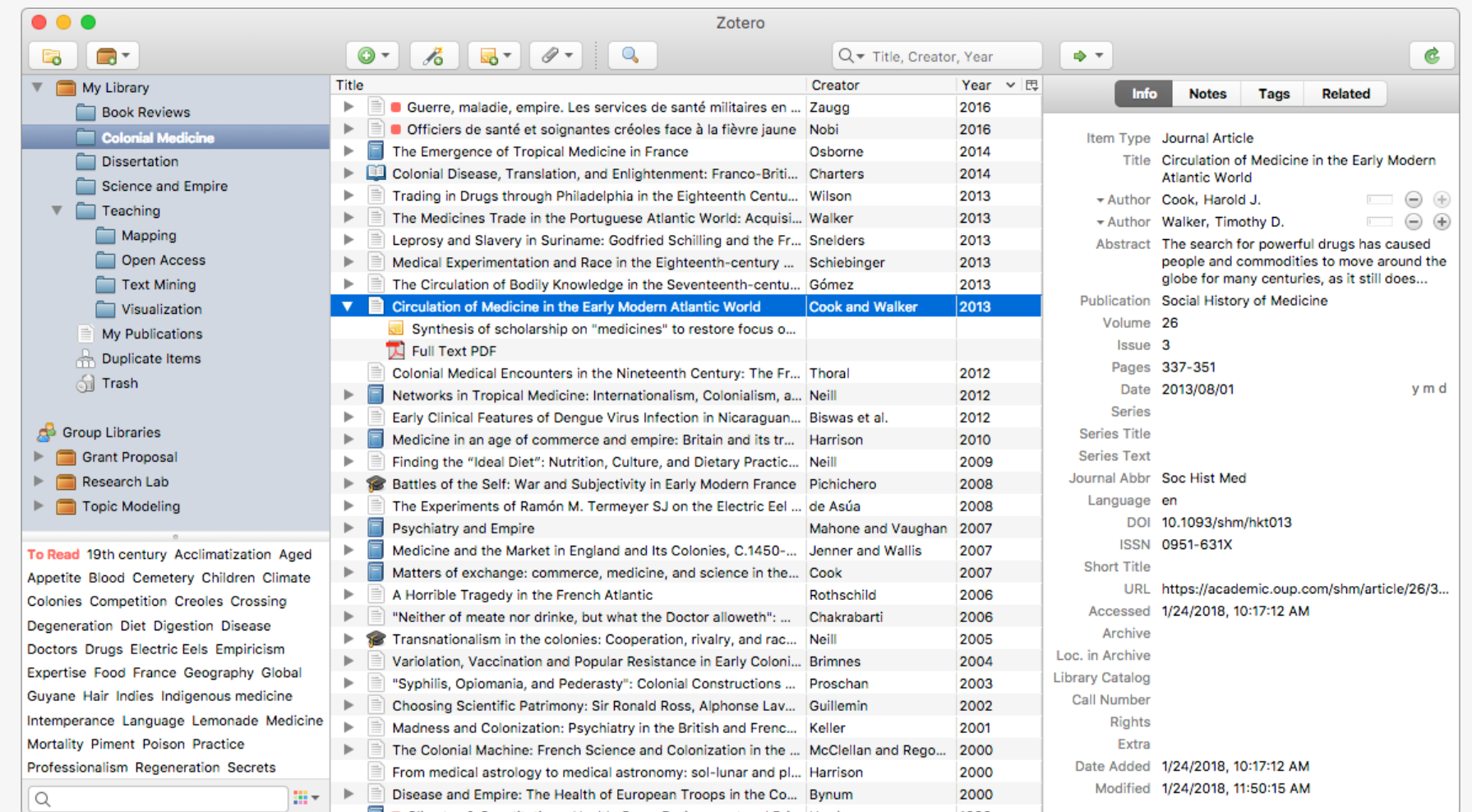- textual data is challenging

- machine-readibility

The Zen of Organization

Empirical Research means Organizing

- scripts as documentation

- efficient automation
  "don't repeat yourself"

- less error-prone

- reproducible

Organizing Literature

- manage literature in one place
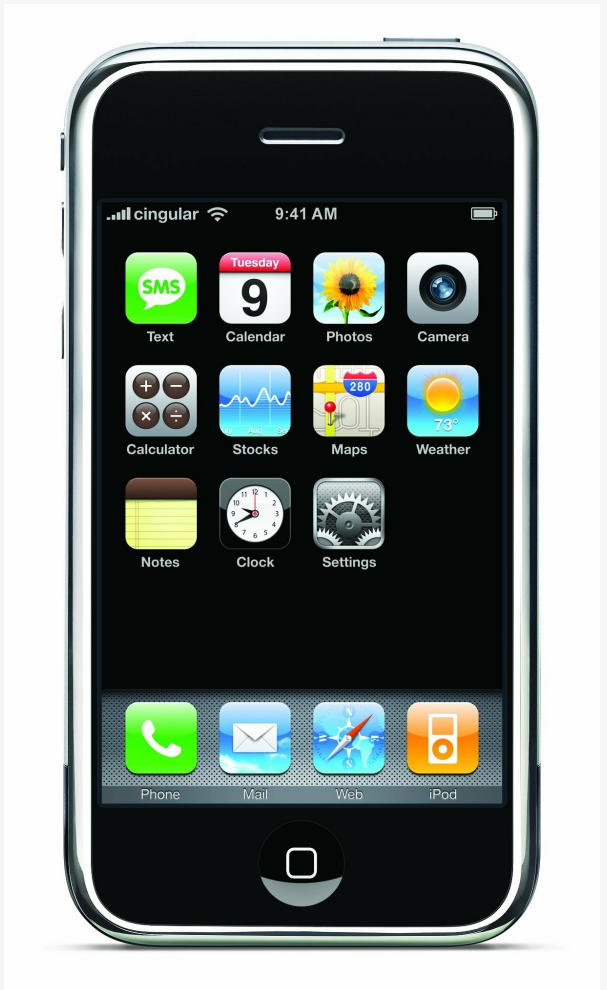
- generating bibliographies

- collect with a click

Computers

Two Trends in Computing

- everyday, mobile, cloud-centered, touch/click-based model
    - *single-purpose application*
    - *smartphones*

- engineering, open-source, machine-oriented model
    - *powerful due to modularity*
    - *workflow for data science*
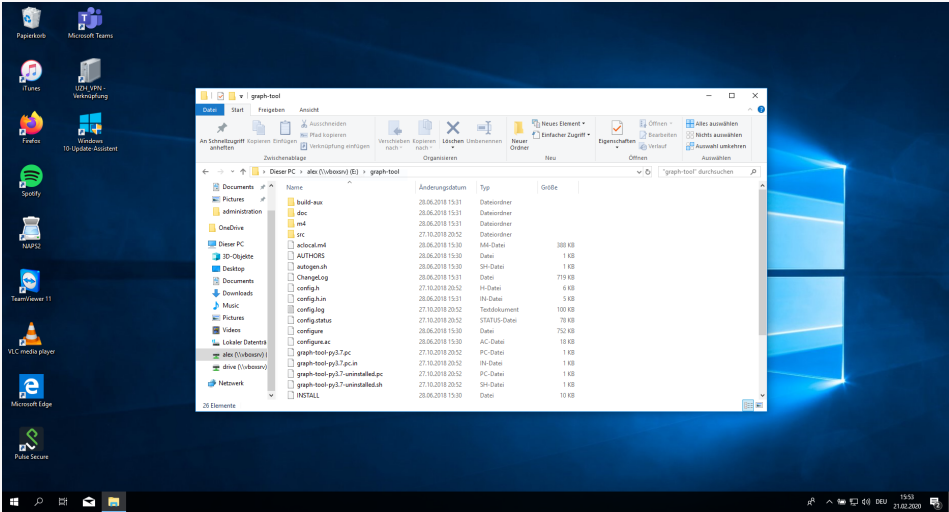
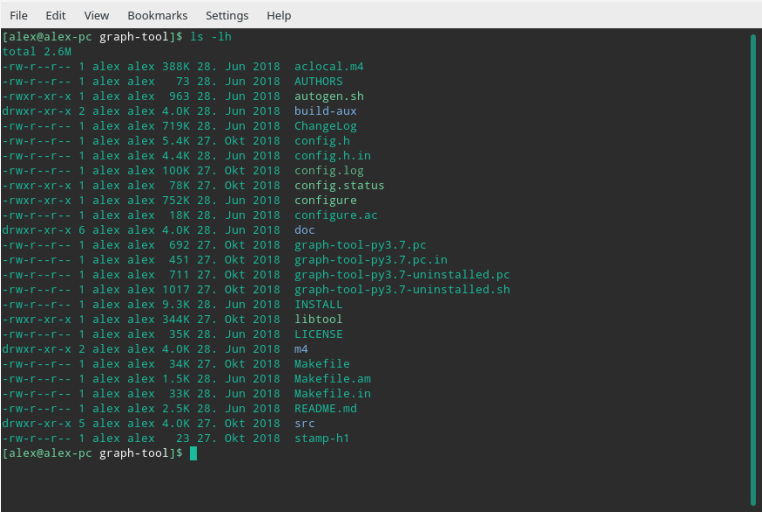*(see also Healy 2019)*

Operating Systems (OS)

- Microsoft Windows

- macOS

- Linux / Unix 🤓

User Interfaces

graphical user interface (GUI)

command line interface (CLI)

"An awesome Programmer saves the World"

Human-Machine Interaction

- text commands instead of clicks

- similar to a 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
    cooking

Package Manager

- automation of software installation

  system: apt, Homebrew
  Python: conda, pip

- software dependencies

- similar to app stores

The Spirit of Open-Source

" *«Standing on the shoulders of giants.»* "

- free + open software

- cross-platform portability

- collaboration
  share + reuse

- helpful community

Resources

- where you get answers to technical questions
    Stack Overflow

- where you find open software
    GitHub

Remember 👍

«Learning by doing, doing by googleing.»

Set up your System

Backup

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

Installation: Setting up your Development Environment

Installation

- Anaconda Platform with Python 3.7
  packages for NLP + visualization

- Tesseract

- various Bash tools

Follow the installation guide for your OS.

Relax.
It takes a while.

Spyder IDE

- text editor

  search/replace, casing

- integrated development environment (IDE)

  debugging, interactive development

  similar to RStudio

First steps in Python

Learn about Python in the interactive tutorials

- Python Principles

- LearnPython

Readings

optional: pimp your workflow

Healy, Kieran. 2019. "The Plain Person's Guide to Plain Text Social Science." online.

References

Healy, Kieran. 2019. "The Plain Person's Guide to Plain Text Social Science."
https://kieranhealy.org/publications/plain-person-text/.