The ABC of Computational Text Analysis

#1 Introduction +
Where is the digital revolution?

Alex Flückiger
27 February 2020

Faculty of Humanities and Social Sciences
University of Lucerne

Outline

1. digital revolution or hype?
2. about us
3. goals of this course

Today's Lecture

The world has changed, hasn't it?

An Era of Big Data + AI

Group Discussion

What makes a computer looking intelligent?

AI is a moving target with respect to …

- *human capabilities*
- *technological abilities*

Mimicking Human Intelligence

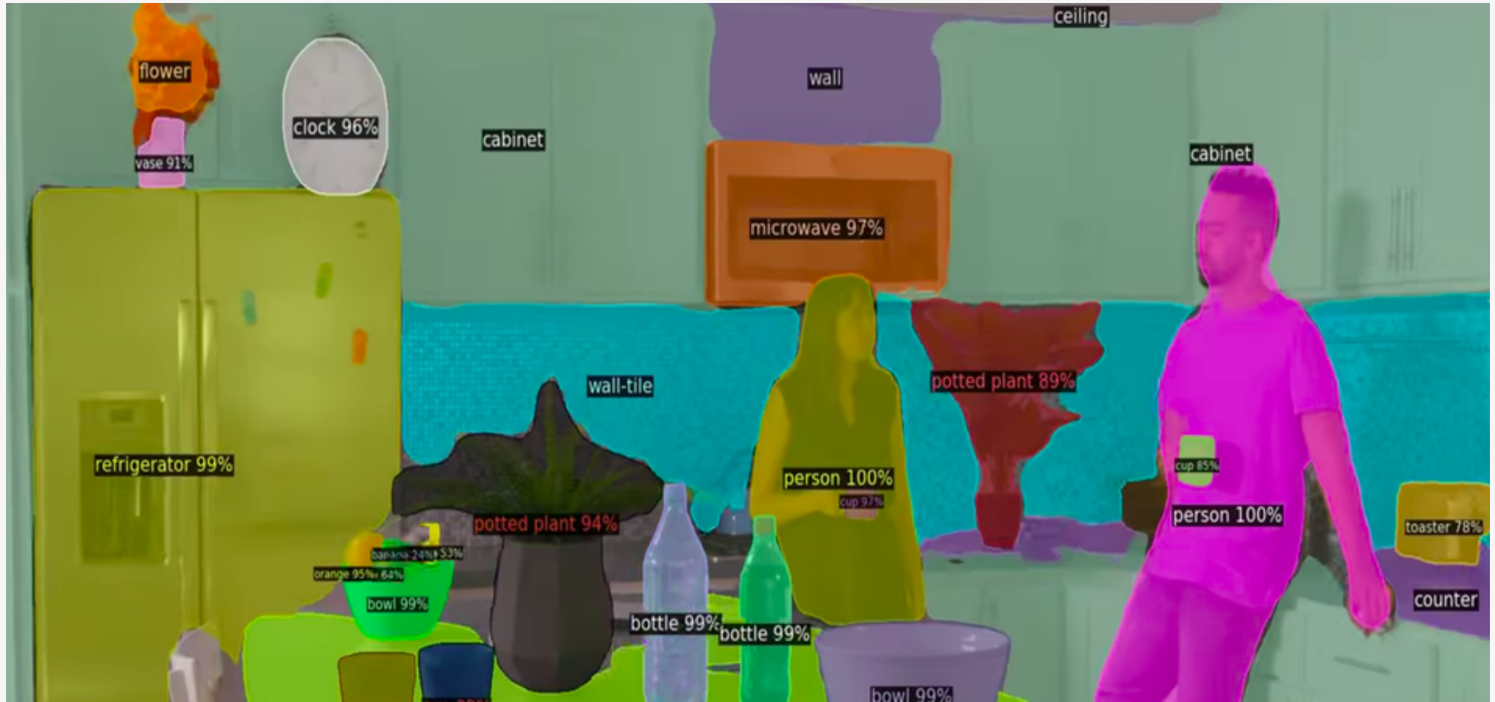
from static machines to more flexible devices

- mimicking intelligent behavior

  speaking + seeing

- a sense of contextual perception
- many degrees of freedom

Seeing like a Human



Speaking like a Human?

Meena Conversation 4	
Human:	*Hi!*
Meena:	*Hey! How's it going?*
Human:	*How's what going?*
Meena:	*Pretty good. How's your day been?*
Human:	*Excellent. I spent most of trying to learn to surf.*
Meena:	*Sweet! What kind of surfing?*
Human:	*The kind where I try to catch a wave and fail*
Meena:	*Haha. Do you have a surfboard?*
Human:	*I borrowed a friend's, to see if I like it. I'm still a beginner.*
Meena:	*Ah. I'm sure you'll get the hang of it.*
Human:	*Appreciate the vote of confidence :) Have you ever been surfing?*
Meena:	*I have! It's actually a ton of fun.*

Beyond Perception

Generated Images by a Neural Network

https://thisxdoesnotexist.com/

Give me *more!*

Demos of *Artificial Intelligence*

Can you disenchant them?

- sentiment analysis
- question answering
- image caption
- text generation
- chat bots
- machine translation

Artificial Intelligence?

Subfields

- Natural Language Processing (NLP)
- Computer Vision (CV)
- Robotics

```
AI = from human_knowledge import solution
```

How does Computer Intelligence work?

- interchangeably (?) used names

  *Artificial Intelligence (AI), Machine Learning (ML), **Deep Learning** (DL)*

- generalize **patterns** from lots of data

  more recycling than genuine intelligence

  theory agnostically

- supervised **training**

  input-output pairs

Why this matters for Social Science

Computational Social Science

data-driven research

- computational social science *(Lazer et al. 2009)*

  Digital Humanities, Computational History, Data Science

- highly interdisciplinary
- early computational history already in 1960s *(Graham, Milligan, and Weingart 2015)*

Group Discussion

What kind of data are there?

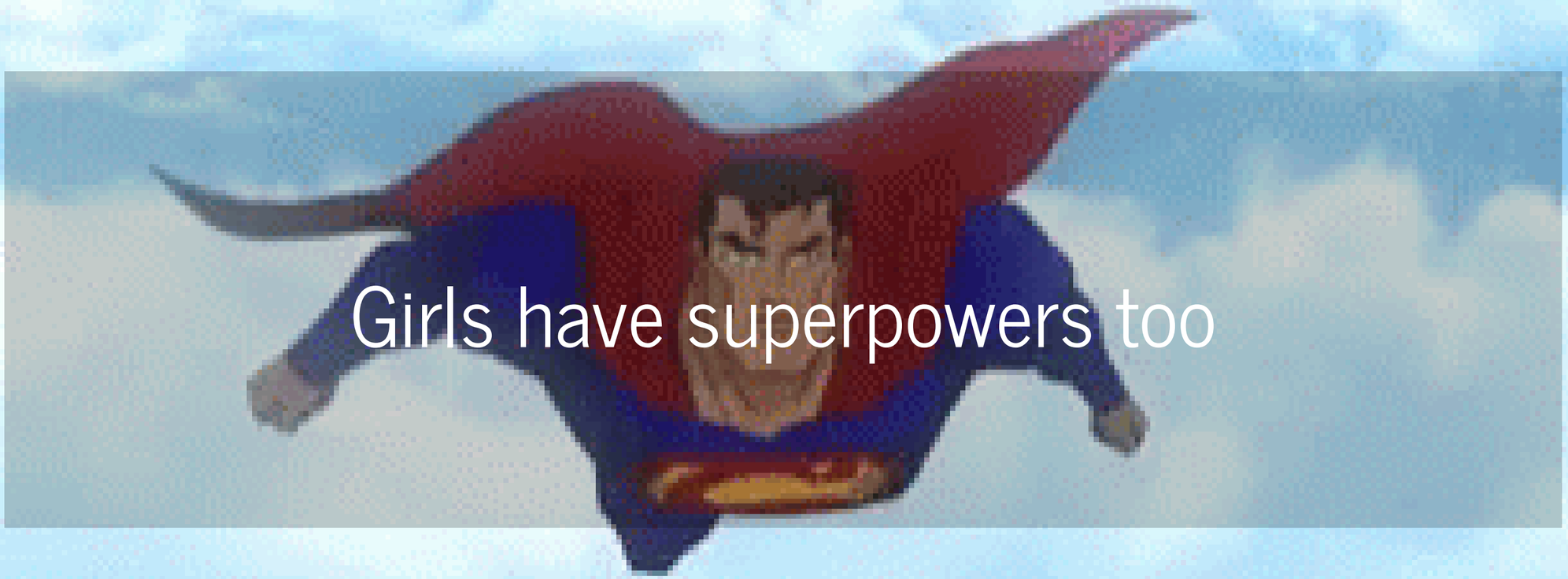
What data is relevant for social science?

- *data as traces of social behaviour*
  tabular, text, image
- *datafication*
  sensors of smartphone, digital communication
- *much of human knowledge compiled as text*

About the Mystery of Coding

coding is like...

- cooking
- superpowers


Girls have superpowers too

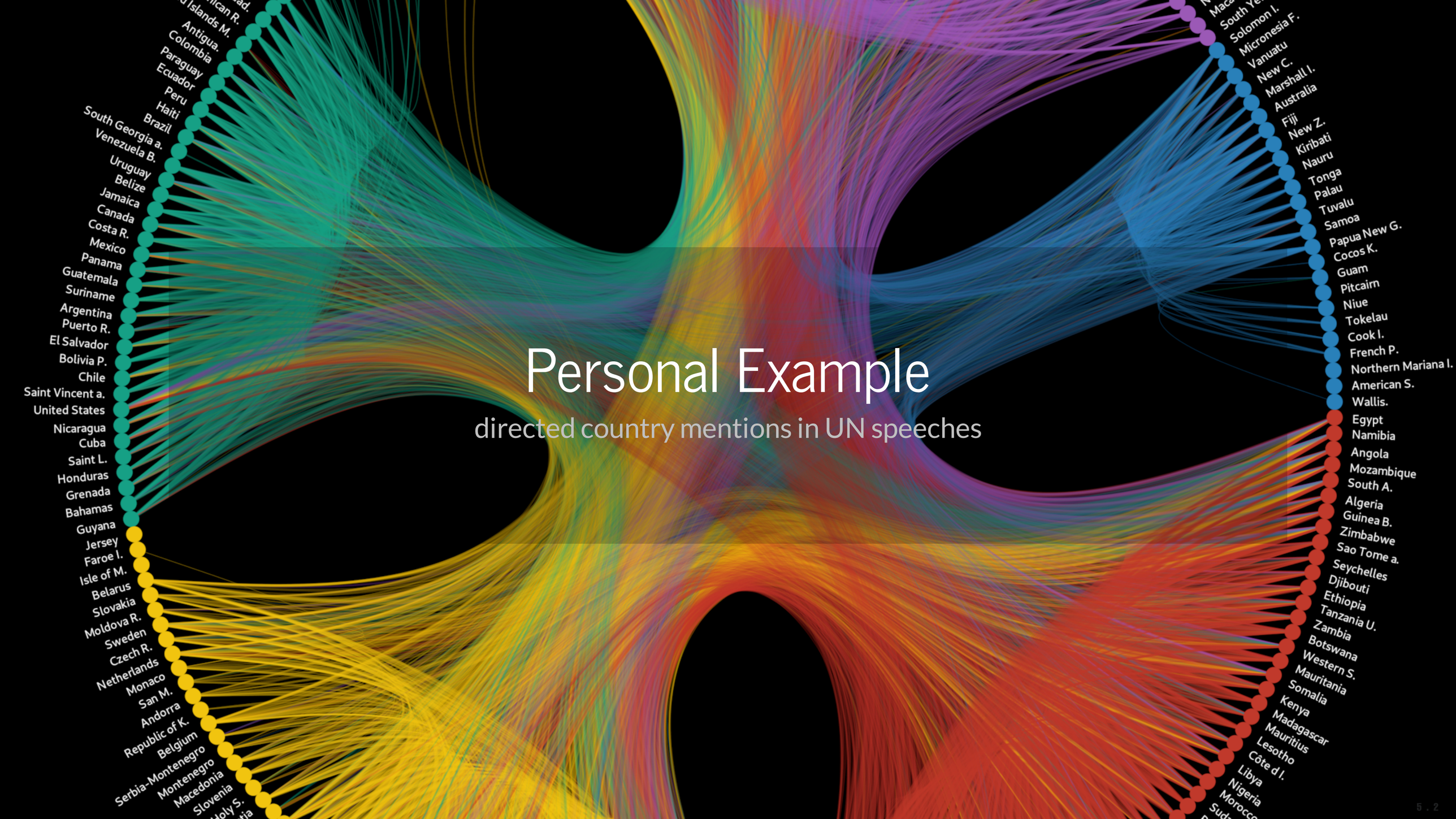
Where the actual Revolution is

Coding is a **superpower**…

- flexible
- reusable
- reproducible
- inspectable
- collaborative

… to tackle complex problems

About us


Personal Example
directed country mentions in UN speeches
Antigua.
Colombia
Paraguay
Ecuador
Peru
Haiti
Brazil
South Georgia a.
Venezuela B.
Uruguay
Belize
Jamaica
Canada
Costa R.
Mexico
Panama
Guatemala
Suriname
Argentina
Puerto R.
El Salvador
Bolivia P.
Chile
Saint Vincent a.
United States
Nicaragua
Cuba
Saint L.
Honduras
Grenada
Bahamas
Guyana
Jersey
Faroe I.
Isle of M.
Belarus
Slovakia
Moldova R.
Sweden
Czech R.
Netherlands
Monaco
San M.
Andorra
Republic of K.
Belgium
Serbia-Montenegro
Montenegro
Macedonia
Slovenia
Solomon I.
Micronesia F.
Vanuatu
New C.
Marshall I.
Australia
Fiji
New Z.
Kiribati
Nauru
Tonga
Palau
Tuvalu
Samoa
Papua New G.
Cocos K.
Guam
Pitcairn
Niue
Tokelau
Cook I.
French P.
Northern Mariana I.
American S.
Wallis.
Egypt
Namibia
Angola
Mozambique
South A.
Algeria
Guinea B.
Zimbabwe
Sao Tome a.
Seychelles
Djibouti
Ethiopia
Tanzania U.
Zambia
Botswana
Western S.
Mauritania
Somalia
Kenya
Madagascar
Mauritius
Lesotho
Côte d I.
Libya
Nigeria

Goals of this Course

What you learn

- **analyze**, interpret, and visualize **texts** with **computational** methods
- **digital literacy** + scholarship
- problem-**solving** capacity

Learnings from previous Courses

- too much content, too little **practice**
- programming can be overwhelming
- **learning** by **doing**, doing by **googling**

Levels of Proficiency

1. awareness of today's computational potential
2. analyzing existing datasets
3. creating + analyzing new datasets
4. applying advanced machine learning

What I teach

- solid **computational** foundation

  command line + python

- **critical perspective** on technology
- lecture-style introductions
- hands-on coding sessions
- discussions + experiments in groups

Topics

techniques

- text processing
- extract information
- basic visualization
- optical character recognition (OCR)
- scraping files

data

- existing resources
- creating new resources

 inputs are more than welcome

TL;DR

You will be tech-savvy…

…yet no programmer applying fancy machine learning

Provisional Schedule

Date	Topic
25 February 2021	Introduction + Where is the digital revolution?
04 March 2021	Text as Data
11 March 2021	Setting up your Development Environment
18 March 2021	Introduction to the Command-line
25 March 2021	Basic NLP with the Command-line
01 April 2021	Learning Regular Expressions
08 April 2021	*no lecture (Osterpause)*
15 April 2021	Advanced RegEx + Data Sources
22 April 2021	Creating new Data Sets + Ethics
29 April 2021	Introduction to Python
06 May 2021	NLP with Python
13 May 2021	*no lecture (Christi Himmelfahrt)*
20 May 2021	NLP with Python + Working Session
27 May 2021	Mini-Project Presentations + concluding Discussion
03 June 2021	*no lecture (Fronleichnam)*

Requirements

- no technical skills required
  self-contained course
- laptop (macOS, Win10, Linux) 
  update OS
  free up at least 15GB storage
  backup

Grading

- 3 exercises during semester
  no grades (pass/fail)
- mini-project with presentation
  backup claims with numbers
  work in teams
  data of your interest
- optional: writing a seminar paper
  in cooperation with Prof. Sophie Mützel

Organization

Seminar on Thursday from 2.15pm - 4.00pm

slides + data on github: KED2021

readings on OLAT: link to course

questions to: alex.flueckiger@doz.unilu.ch

Who are you?

Please fill out this questionnaire


Questions?

Reading

Required

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King,

Michael Macy, Deb Roy, and Marshall Van Alstyne. 2009. “Computational Social Science.” *Science* 323(5915):721–23.

(via OLAT)

Optional

Graham, Shawn, Ian Milligan, and Scott Weingart. 2015. *Exploring Big Historical Data: The Historian’s Macroscope*. Open Draft Version. Under contract with Imperial College Press.

online

References

Graham, Shawn, Ian Milligan, and Scott Weingart. 2015. *Exploring Big Historical Data: The Historian's Macroscope*. Open Draft Version. Under contract with Imperial College Press. http://themacroscope.org.

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.