The ABC of computational Text Analysis

01: Introduction + Where is the digital revolution?

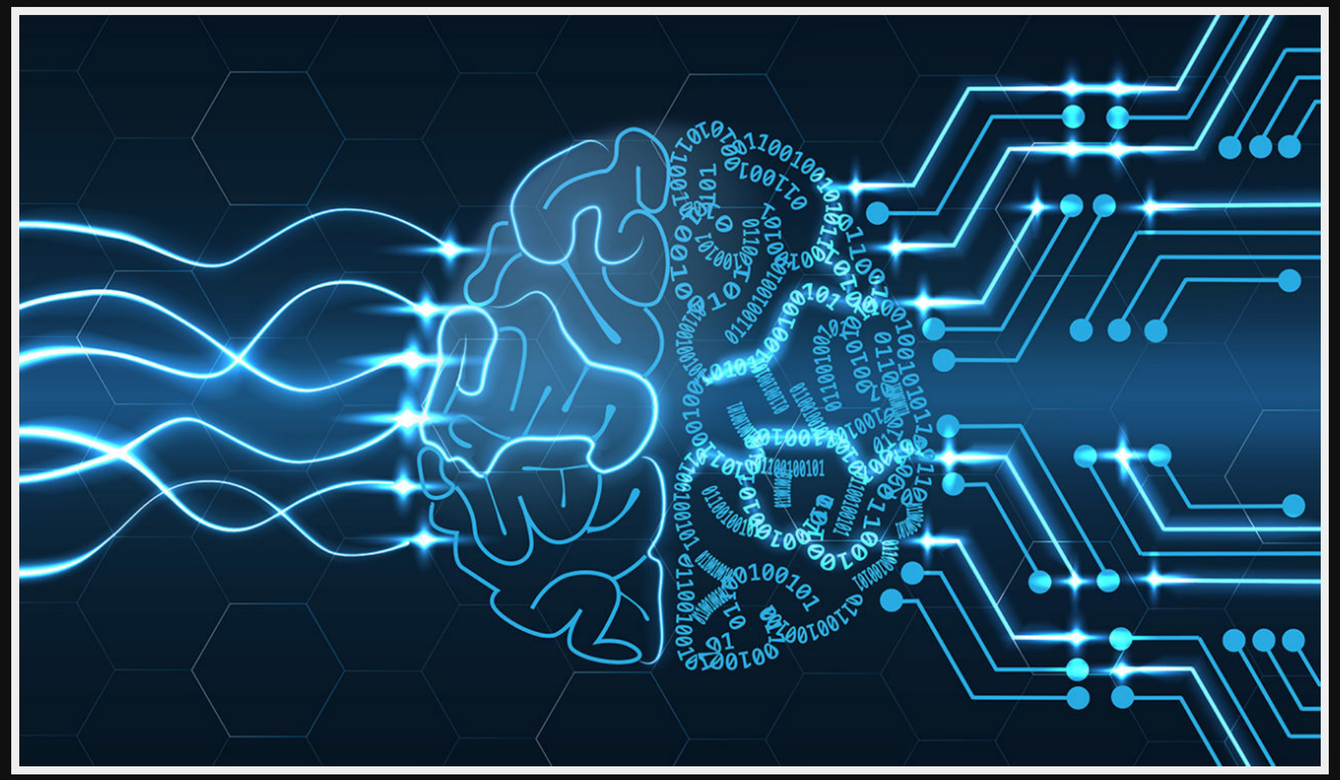
Alex Flückiger
27 February 2020

Outline

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

Today's Lecture

The world has changed, hasn’t it?



symbol of artificial intelligence (HWZ)

An Era of Big Data + AI

Group Discussion

- What makes a computer looking intelligent?

> AI not absolute, rather relative 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

image segmentation with Facebook's Detectron2

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.*
Human:	*Did you find it hard at first?*
Meena:	*A little bit, yes. But once I got the hang of it, it was really fun.*
Human:	*That's good to know. Do you go surfing regularly?*
Meena:	*Yup. I live in southern Arizona, so there's plenty of surfing to be had.*

Chatting with Google's Meena

Not really, Arizona is not by the sea.

Demos of *Artificial Intelligence*

Can you disenchant them?

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

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, texts, images
- 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

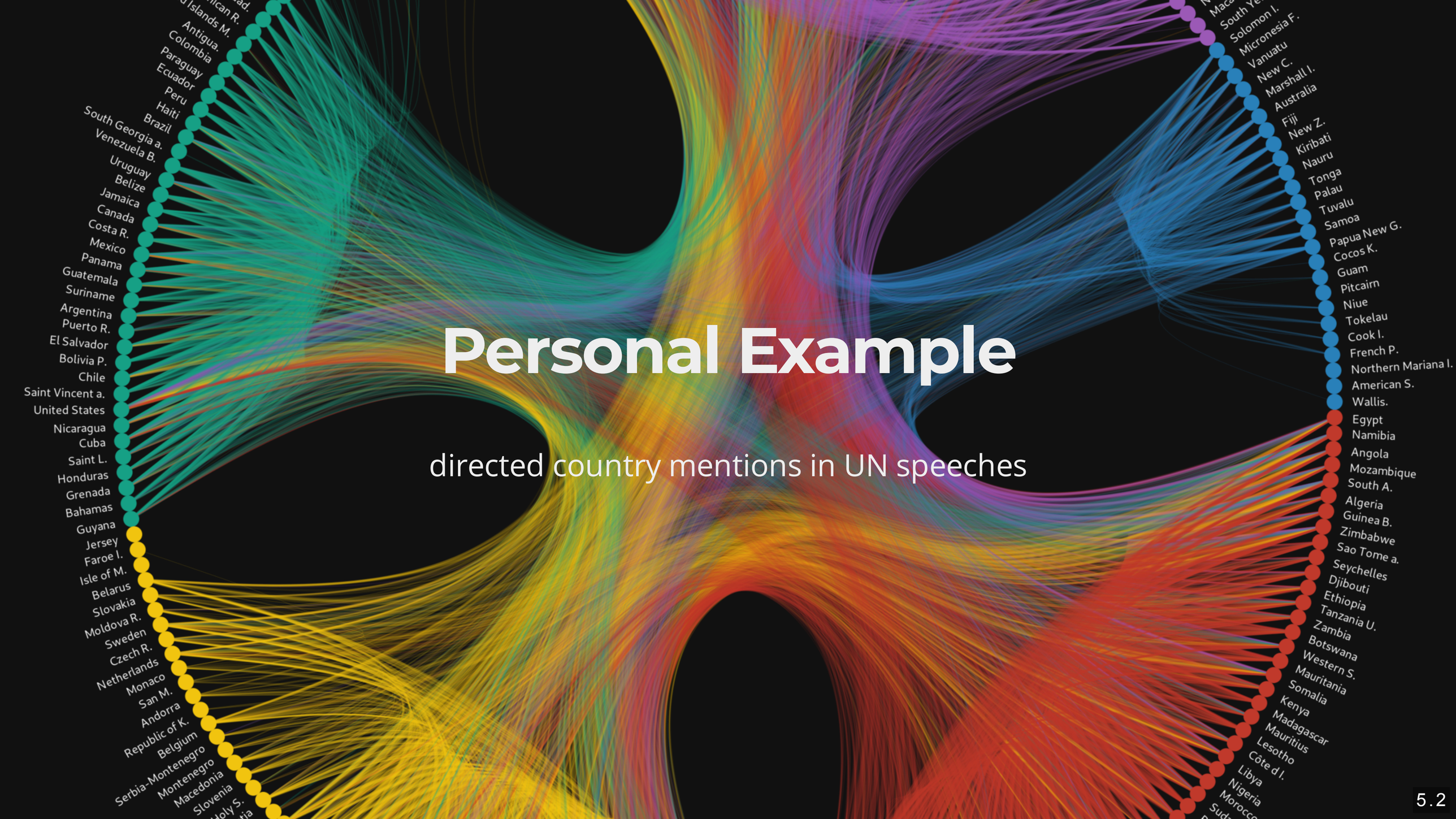
Where the actual Revolution is

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

- analyzing texts using a computer
- digital scholarship
- problem solving capacity

Learnings from previous Courses

- too much content, too less practicing
- 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
- lecture-style introductions
- practical 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
27 February 2020	Introduction + Where is the digital revolution?
05 March 2020	Text as Data
12 March 2020	Setting up your Development Environment
19 March 2020	Introduction into Command-line
26 March 2020	Basic NLP with Command-line
02 April 2020	Learning Regular Expressions
09 April 2020	Data Sources + Ethics
16 April 2020	*no lecture (Osterpause)*
23 April 2020	Creating new Data Sets
30 April 2020	Introduction to Python
07 May 2020	NLP with Python
14 May 2020	NLP with Python + Working Session
21 May 2020	*no lecture (Christi Himmelfahrt)*
28 May 2020	Mini-Project Presentations + concluding Discussion

Requirements

- no technical skills required
- laptop (macOS, Win10, Linux)

  update OS
  free up at least 15GB storage
  backup

Grading

- no grades (pass/fail)
  - *3 exercises during semester*
- mini-project with presentation
  - *complement 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: KED2020

readings on OLAT: link to course

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

Questionnaire

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.

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.