

Text Analytics – Week 3

Program: *IM-KMBI*

Teacher: *R. Henriques, F. Peleja, J. Jagusch*

Date: *Friday, March 2, 2018*

Let me introduce myself...

- Jan-Benedikt Jagusch
- German
- 2nd year master student in *IM-KMBI*
- B.S. in *IT Management*
- Teaching assistant in *Text Analytics*
and *Decision Support Systems*



Let's start of with a quiz

German: *'Die Kirche im Dorf lassen'*

English literal: *'To leave the church in the village'*



Let's start of with a quiz

German: *'Die Kirche im Dorf lassen'*

English literal: *'To leave the church in the village'*

English equivalent: *'To not get carried away'*

"You won't get a Nobel prize for your master thesis. Leave the church in the village!"



Let's start of with a quiz

German: *'Schwein haben'*

English literal: *'To have a pig'*



Let's start of with a quiz

German: *'Schwein haben'*

English literal: *'To have a pig'*

English equivalent: *'To have a stroke of luck'*

"The car almost hit you. You had a pig!"



Let's start of with a quiz

German: *'Seinen Senf dazu geben'*

English literal: *'To add their mustard'*



Let's start of with a quiz

German: *'Seinen Senf dazu geben'*

English literal: *'To add their mustard'*

English equivalent: *'To express your opinion.'*

"Peter can't stop arguing. He always adds his mustard."



And now, the agenda...

- Introduction to our text analytics environment
 1. GitHub
 2. Anaconda
 3. Jupyter Notebook
- A crash-course in Python
 1. Data structures and types
 2. Code flow constructs
 3. Functions
 4. Classes
- Mini-project: *Making sense of Trump's tweets*

GitHub

- VCS: Version Control System
- Collaboration with other developers on software projects
- Central code repository
- Project development history
- Your portfolio as a data scientist

GitHub

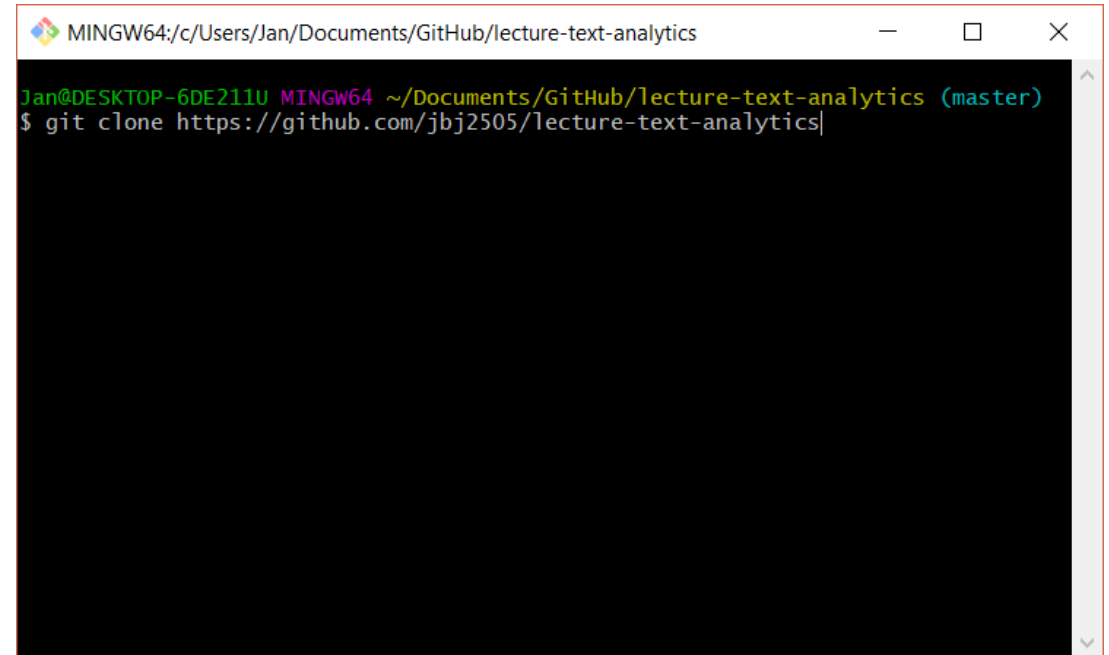


GitHub

- Our GitHub for this course:

<https://github.com/jbj2505/text-analytics-lecture>

- Interactive [GitHub tutorial](#)
- GitHub [cheat sheet](#)

A screenshot of a terminal window titled "MINGW64:/c/Users/Jan/Documents/GitHub/lecture-text-analytics". The terminal shows a prompt "Jan@DESKTOP-6DE211U MINGW64 ~/Documents/GitHub/lecture-text-analytics (master)" followed by the command "\$ git clone https://github.com/jbj2505/lecture-text-analytics|". The terminal background is black with green and yellow text. There are scrollbars on the right side of the terminal window.

```
MINGW64:/c/Users/Jan/Documents/GitHub/lecture-text-analytics
Jan@DESKTOP-6DE211U MINGW64 ~/Documents/GitHub/lecture-text-analytics (master)
$ git clone https://github.com/jbj2505/lecture-text-analytics|
```

Anaconda

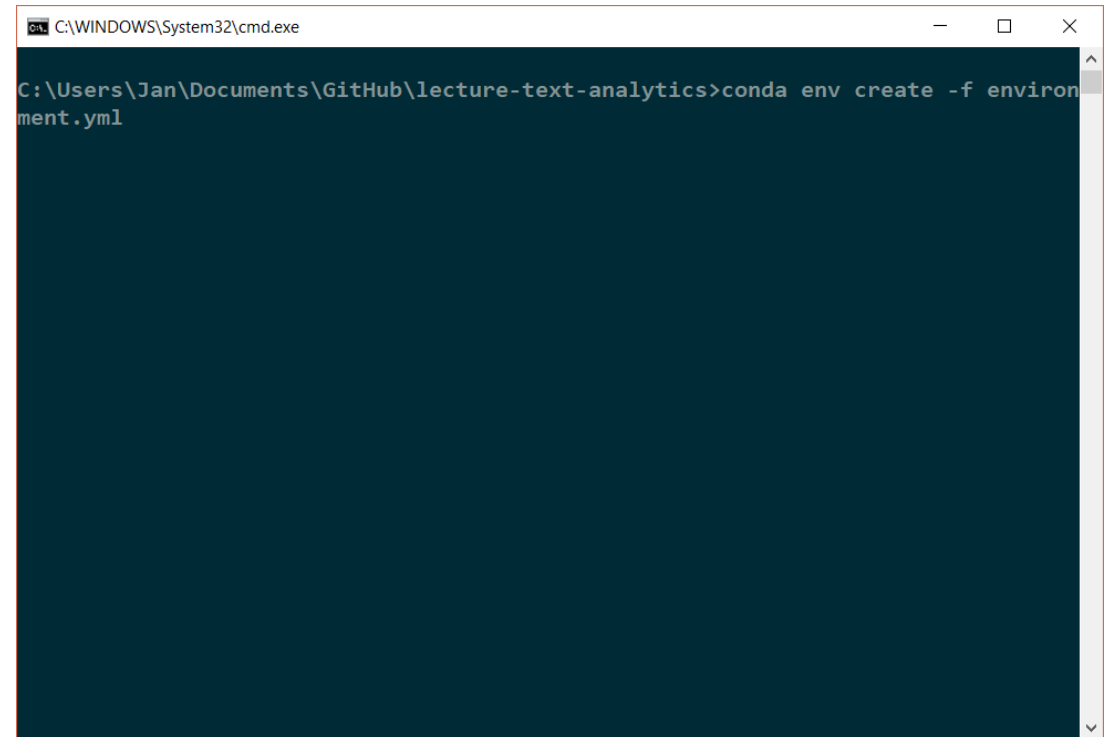
- Data science toolkit for Python
- Conda: *Virtual Environment Manager*
- Data science packages
- Low on disk space? [Miniconda](#)



ANACONDA
Powered by Continuum Analytics®

Anaconda

- Conda [tutorial](#)
- Conda [cheat sheet](#)



A screenshot of a Windows command prompt window. The title bar at the top reads "C:\WINDOWS\System32\cmd.exe". The command prompt shows the following command being entered: `C:\Users\Jan\Documents\GitHub\lecture-text-analytics>conda env create -f environment.yml`. The command is partially visible, with "environment.yml" on the second line. The background of the command prompt is dark blue.

Jupyter Notebook

- Markdown based development environment
- Used for coding and documentation



A crash-course in Python...

- Data extracted from [Trump Twitter Archive](#)
- Data from 01.01.2013 until 31.12.2017
- 28,282 *tweets* in our collection

