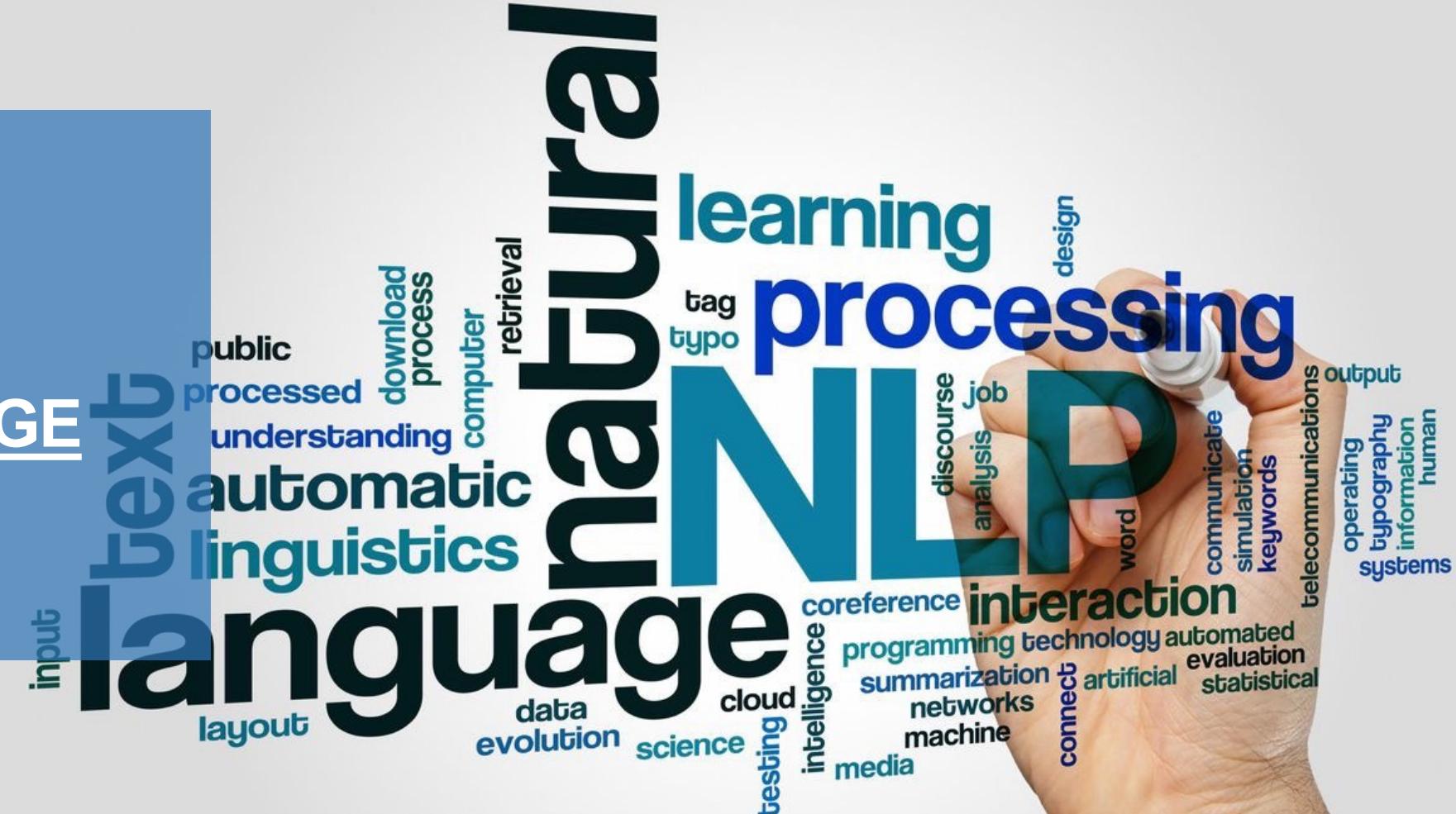




INTRODUCTION TO NATURAL LANGUAGE PROCESSING



LECTURER AND TEACHING ASSISTANT



Lucie Flek

flek@bit.uni-bonn.de



Vahid Sadiri Javadi

vahidsj@bit.uni-bonn.de

Lectures: **Mondays** 10:30 – 12.00 (B-IT-Max 0.109) ([Zoom Link](#))

Exercises: **Mondays** 16:00 - 18:00 (B-IT-Max 0.109) ([Zoom Link](#))

[eCampus Course](#)

ABOUT OUR GROUP + ME

Current CAISA Team



Lucie Flek



Akbar Karimi



Charlie Welch



Allison Lahnala



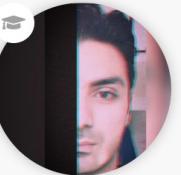
Joan Plepi



Lea Fischbach



Shaina Ashraf



Vahid Sadiri Javadi



Atefeh Soltanifard



Béla Neuendorf



Pepper

Conversational AI and Social Analytics (CAISA) Lab

- Natural Language Processing
- **Conversational AI**
- User Modeling and Computational Social Science
- Machine Learning for NLP
- Privacy and Fairness in Natural Language Applications

<https://caisa-lab.github.io>

Today, we will talk
about:

- Survey: [Link](#)
- Organization
- Introduction to Python
- Software Setup
- Jupyter Notebook
- Recap & Questions



COURSE ORGANIZATION

Content of Course:

Holidays:

- 10.04
- 01.05
- 29.05



- 03.04.2023 | Introduction & Python basics

Feature Engineering:

- 17.04.2023 | Word operations & Feature extraction using Pandas, Sklearn
- 24.04.2023 | Linear classification using TF - IDF

Language Processing:

- 08.05.2023 | Word embeddings using spaCy
- 15.05.2022 | Q & A: PF + PS
- 22.05.2023 | POS tagging & HMMs
- 05.06.2023 | Transformers and Generative Models I
- 12.06.2023 | Transformers and Generative Models II
- 19.06.2023 | Project development (supervision by appointment)
- 26.06.2023 | Project development (supervision by appointment)
- 03.07.2023 | Project development (supervision by appointment)

10.07.2023 | PROJECT PRESENTATIONS (PP)



- **GRADING:**
- **Exam (60%)**
- **Project (40%) >> PF (Problem Formulation): 10%**
PS (Problem Solving): 25%
PP (Project Presentation): 5%

- **03.04.2023 | Introduction & Python basics**

Feature Engineering:

- 17.04.2023 | Word operations & Feature extraction using Pandas, Sklearn
- 24.04.2023 | Linear classification using TF - IDF

Language Processing:

- 08.05.2023 | Word embeddings using spaCy
- **15.05.2022 | Q & A: PF + PS**
- 22.05.2023 | POS tagging & HMMs
- 05.06.2023 | Transformers and Generative Models I
- 12.06.2023 | Transformers and Generative Models II
- 19.06.2023 | Project development (supervision by appointment)
- 26.06.2023 | Project development (supervision by appointment)
- 03.07.2023 | Project development (supervision by appointment)

10.07.2023 | PROJECT PRESENTATIONS (PP)

■ GRADING:

- Exam (60%)
- Project (40%) >> PF: 10% PS: 25% PP: 5%

■ PROJECT:

- You will experiment with NLP techniques using **Python**
- You can work in a **group** (3 - 5)
- You should choose a dataset, formulate a real-world problem (PF) and try to solve it (PS)
- Teams write a 3 - 5 page summary about their project and present the results with a poster (PP)
- We will provide you with several datasets to work on one task. (Any suggestions of yours are welcome!)
- You will find the points you need to explain in the PF on eCampus soon!

■ ASSIGNMENTS:

- Submitted assignments will be considered as **bonus points** for the students who want to write their master thesis with us!
- During the semester, you will receive 6 assignments.
- Assignments will be shared on eCampus and should be submitted either **individually or in a group of two students!**
- Please comment your code.
- All assignments will use Python as programming language.

| Assignment | Assignment Submission |
|------------|-----------------------|
| 1 | 23.04.2023 |
| 2 | 30.04.2023 |
| 3 | 14.05.2023 |
| 4 | 28.05.2023 |
| 5 | 12.06.2023 |
| 6 | 19.06.2023 |



- Submission of Assignments:
- Sundays at 23:59



■ DEADLINES:

- Submission of team members:

- Monday, April 17th, 23:59

Subject: ITNLP - SS2023 - <Matr. Nr.>

Team Speaker: <Name>, <Matr. Nr.>, <Mail Addr.>

Team Members: <Name>, <Matr. Nr.>, <Mail Addr.>
<Name>, <Matr. Nr.>, <Mail Addr.>

* Lonely o. shy o. independent students:

Subject: ITNLP - SS2023 - LSI - <Matr. Nr.>

E-Mail: itnlp.uni.bonn@gmail.com

- Submission of Problem Formulation (PF): (PDF)

- Sunday, May 14th, 23:59

- Submission of Project Presentation (PP): (Poster)

- Sunday, July 9th, 23:59

- Submission of Final Project Report + Code

- Sunday, August 7th, 23:59

eCampus: [INTRODUCTION TO NLP](#)

THEESIS TOPICS

Open Theses Topics (Master/Bachelor/Project)



[Data Quality Improvement; Evaluate Arguments for Generating Opinionated Sales Conversations](#)
Argumentation and debating are the process of forming reasons that humans engage in. Sales negotiation is one of the conversational activities in which a buyer and a seller communicate reasons to arrive at a satisfactory selection



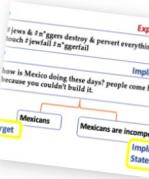
[A poor man's opinion detection tool, training a model with a handful of data](#)

Opinion detection aims to detect an author's view towards a certain topic and has become a key component in everyday applications such as fake news detection and argumentation. While state-of-the-art deep learning models are



[Investigating the level of stubbornness regarding sociopolitical views in social media](#)

While social media platforms help to connect people worldwide and give access to enormous amounts of diverse information, they also foster an environment that promotes polarization. This occurs due to the fact that users



[Implicit Hate Speech Detection](#)

Implicit hate speech is defined by coded or indirect language that disparages a person or group on the basis of protected characteristics like race, gender, and cultural identity. Compared to explicit hate speech detection, implicit hate

[Exploring user context for detecting intended and perceived sarcasm](#)

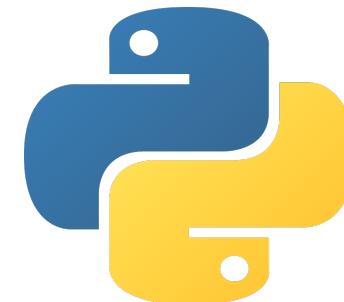
Sarcasm is a form of irony that occurs when there is a discrepancy between the literal meaning of an utterance and its intended meaning. Existing sarcasm detection systems focus on exploiting linguistic markers, context, or userlevel

<https://caisa-lab.github.io/theses/>



**NOW, LET'S TALK
ABOUT PYTHON!**

A BIT OF CONTEXT ...



"Programming languages are how programmers express and communicate ideas — and the audience for those ideas is other programmers, not computers."

-Guido van Rossum

- Created by Guido van Rossum
 - Dutch Programmer
 - Since November 2020 at Microsoft
 - Check [his blog](#) or [this interview](#) if you want to know more
- Described by its creator as "a programming language created by a community"
- The first version of Google was written in Python and Java

ABOUT PYTHON ...

Why do we want to use Python?

Because it...

- has a simple easy-to-use syntax
- is highly readable
- provides powerful debugging
- doesn't require declaration variables
- is very well documented

Note: This is mainly advertising from python.org



IT'S TIME TO MOVE TO PYTHON

FIRST, INSTALLING PYTHON



```
def install_python:  
    if not Python already installed:  
        Install Python following instructions from  
        python.org LINK  
    return "Installed Python!"
```

| Option 1 | Option 2 |
|------------------------|-------------------------|
| pip install virtualenv | easy_install virtualenv |



Follow these steps in your terminal:

```
> mkdir intro2nlp  
> cd intro2nlp  
> virtualenv --python python<3.7*> nlpnv  
(*Check the installed Python version and change it accordingly.)  
> source nlpnv/bin/activate  
> pip install jupyterlab  
> pip install -r requirements.txt  
> jupyter lab
```

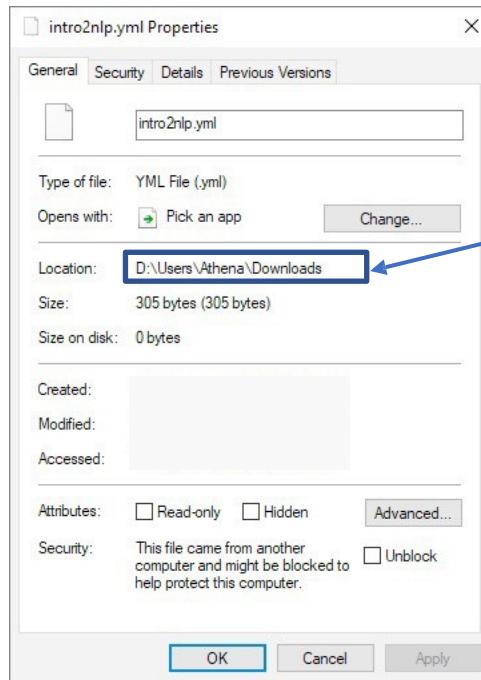
LINKS:

[Virtual Python Environment \(Windows\)](#)

[Jupyter Lab Installation](#)

[Jupyter Notebook Installation](#)

OR CONDA IN WINDOWS



Step 4.

Step 5.

1. Download Miniconda in this [link](#)
2. Find our "*intro2nlp.yml*" file on eCampus
3. Open the Anaconda Prompt on your computer
4. Navigate to the directory where you saved the yml-file (eg: `cd /Downloads/`)
5. Type the following line:
"conda env create -f intro2nlp.yml"
6. Now you're ready to activate your environment with:
"conda activate intronlp"
7. Last step navigate to your class directory and type:
"jupyter notebook"

IF YOU STILL HAVE PROBLEMS WITH INSTALLATION



- You might consider using the notebook directly from [Colaboratory](#).
- It is a [Programming Platform](#) offered by Google. In this case you would need a Google Account.



See you next
Monday!