KI-Campus
Die Lernplattform
für Künstliche Intelligenz

Deutsches
Forschungszentrum
für Künstliche
Intelligenz GmbH

berlin

GEFÖRDERT VOM
Bundesministerium
für Bildung
und Forschung

Natural Language Processing

Salar Mohtaj | DFKI

KI-Campus

Natural Language Processing

- Introduction to Natural Language Processing
- Everyday NLP applications
- Main NLP tasks
- Main approaches in NLP
- Who is this course for
- The Course structure
- NLP terminology

KI-Campus

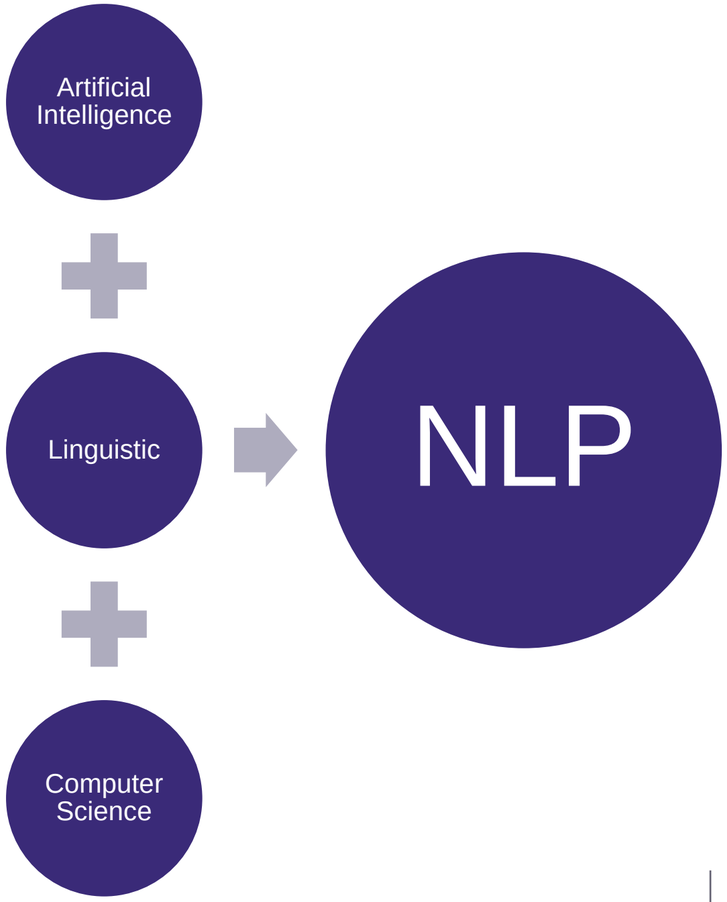
Natural Language Processing

- Introduction to Natural Language Processing
- Everyday NLP applications
- Main NLP tasks
- Main approaches in NLP
- Who is this course for
- The Course structure
- NLP terminology

KI-Campus

Introduction to Natural Language Processing

- Natural language processing (NLP) is a branch of ***artificial intelligence*** that helps computers to understand, interpret and generate human language
- ***Natural language processing*** helps computers communicate with humans in their own language
- Most NLP techniques rely on machine learning to derive ***meaning*** from ***human languages***



KI-Campus

Why is it difficult?

The hammer hit the glass and it broke!

Images from www.creativehomeidea.com and www.moneyversed.com

KI-Campus

Why is it difficult?

I saw someone on the hill with a telescope!

Images from www.thedailychain.com and www.storyblocks.com

Example from www.examples.yourdictionary.com

Why is it difficult?

- Ambiguity in language
  - The rules that dictate the passing of information using natural languages are not easy for computers to understand
  - Sarcastic remark

That's just what I needed today!

Example from www.examples.yourdictionary.com

Why is it difficult?

- Ambiguity in language
  - The rules that dictate the passing of information using natural languages are not easy for computers to understand
  - Sarcastic remark
  - Multi meaning words

She will **park** the car so we can walk in the **park**.

The committee **chair** sat in the center **chair**.

Example from www.spellingcity.com

Why is it difficult?

- Ambiguity in language
  - The rules that dictate the passing of information using natural languages are not easy for computers to understand
  - Sarcastic remark
  - Multi meaning words
- The lexicon of a language is usually enormous
  - Oxford dictionary has ***273,000*** headwords; ***171,476*** of them being in current use
  - An average person has a vocabulary range of about ***20,000*** to ***35,000***

Examples from https://spellingcity.com

KI-Campus

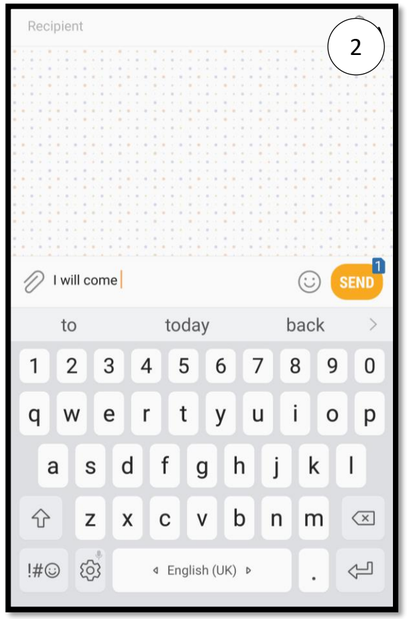
Natural Language Processing

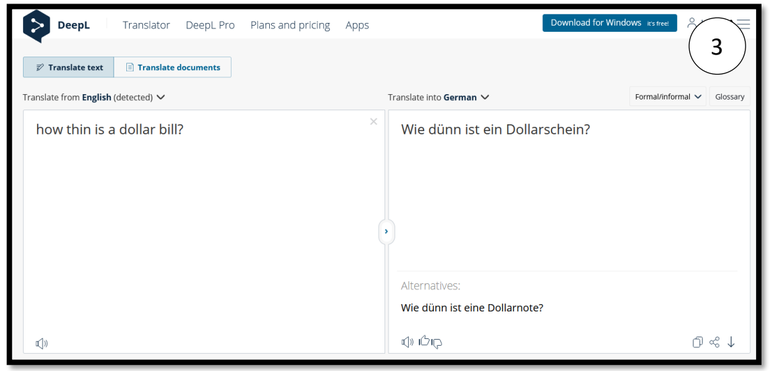
- Introduction to Natural Language Processing
- Everyday NLP applications
- Main NLP tasks
- Main approaches in NLP
- Who is this course for
- The Course structure
- NLP terminology

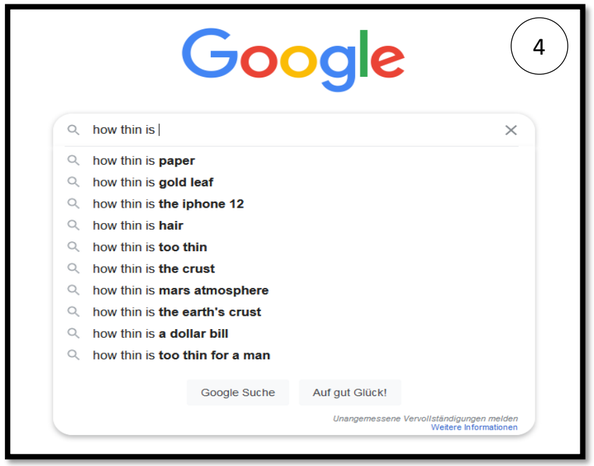
KI-Campus

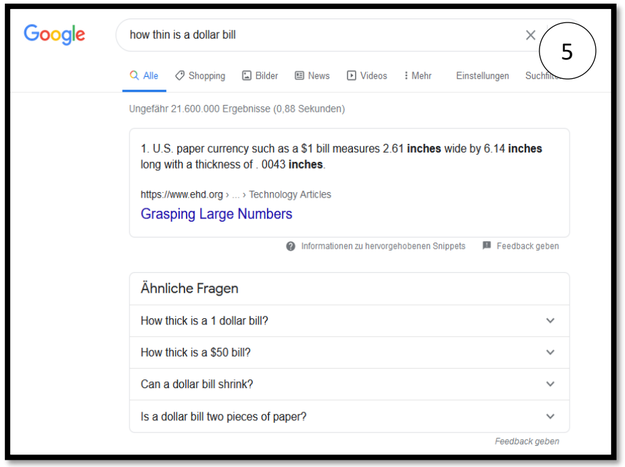
Everyday NLP applications

- Email filters (spam detection) (1)
- Faster typing (2)
- Language translation (3)
- Question answering (4)+(5)
- Smart assistant devices (6)









KI-Campus

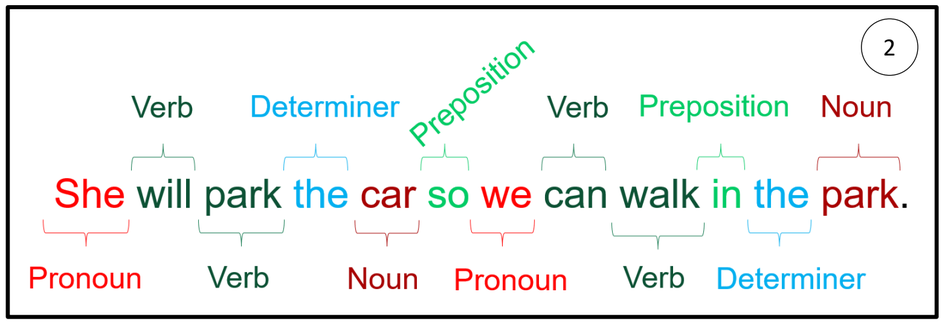
Natural Language Processing

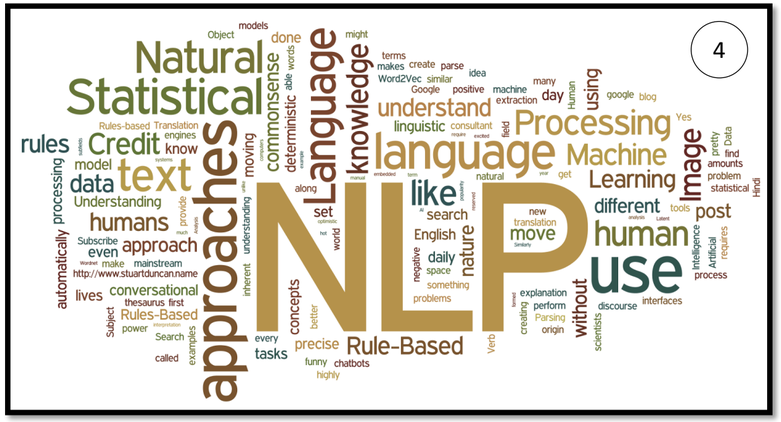
- Introduction to Natural Language Processing
- Everyday NLP applications
- Main NLP tasks
- Main approaches in NLP
- Who is this course for
- The Course structure
- NLP terminology

KI-Campus

The main NLP tasks

- Text classification (1)
- Parts of speech tagging (2)
- Sentiment analysis (3)
- Keyword extraction (4)
- Text similarity (5)

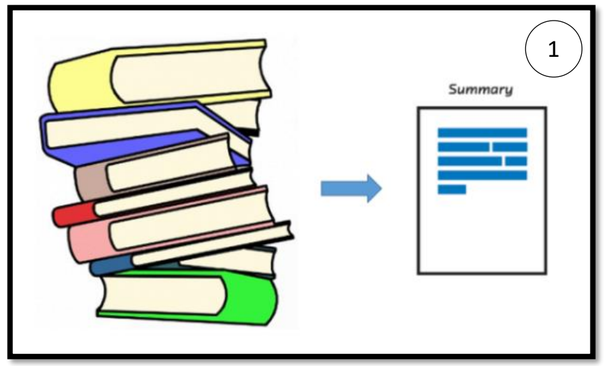


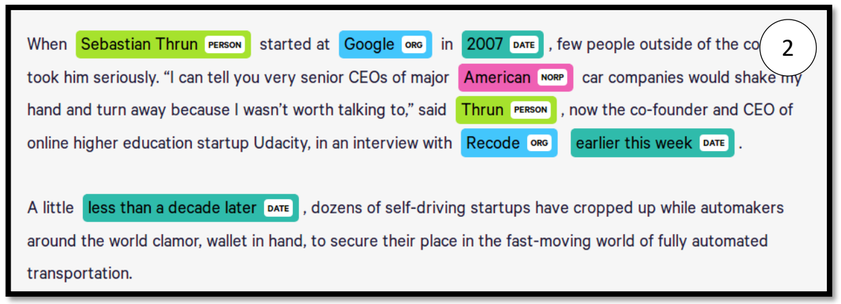


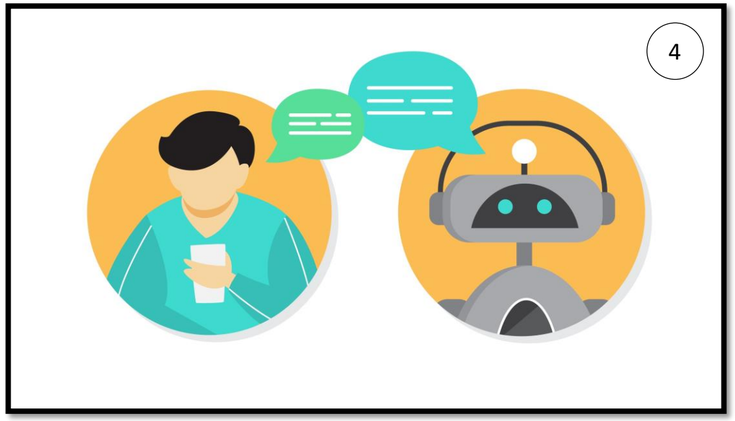
KI-Campus

The main NLP tasks

- Text summarization (1)
- Named entity recognition (NER) (2)
- Machine translation (3)
- Question answering (4)
- Image captioning (5)







KI-Campus

Natural Language Processing

- Introduction to Natural Language Processing
- Everyday NLP applications
- Main NLP tasks
- Main approaches in NLP
- Who is this course for
- The Course structure
- NLP terminology

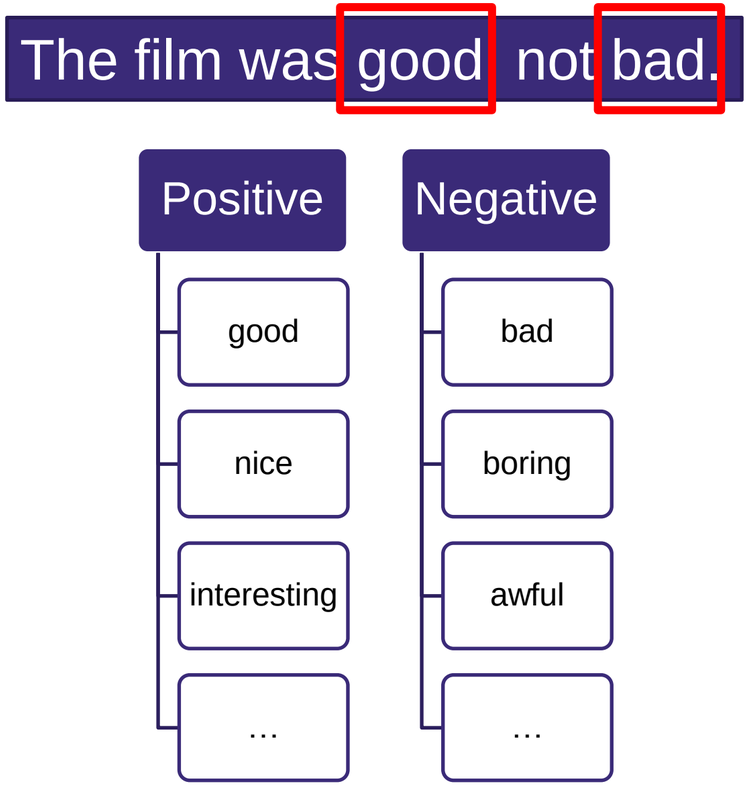
KI-Campus

Main Approaches in NLP

- Rule based approaches
- Classical machine learning
- Deep learning

Main Approaches in NLP

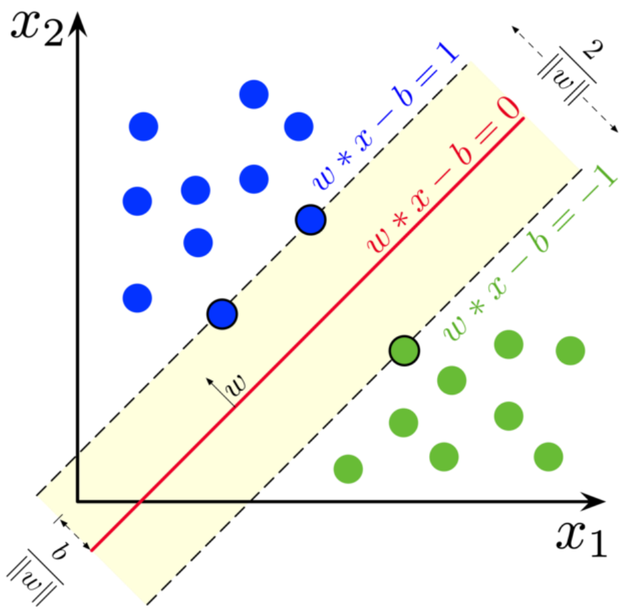
- Rule based approaches
  - Lack of enough accuracy

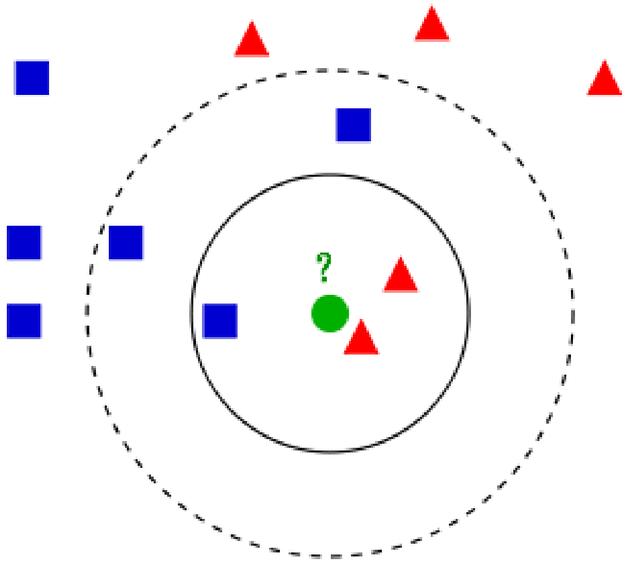


KI-Campus

Main Approaches in NLP

- Rule based approaches
  - Lack of enough accuracy
- Classical machine learning
  - Training data
  - Feature engineering
  - Training a model

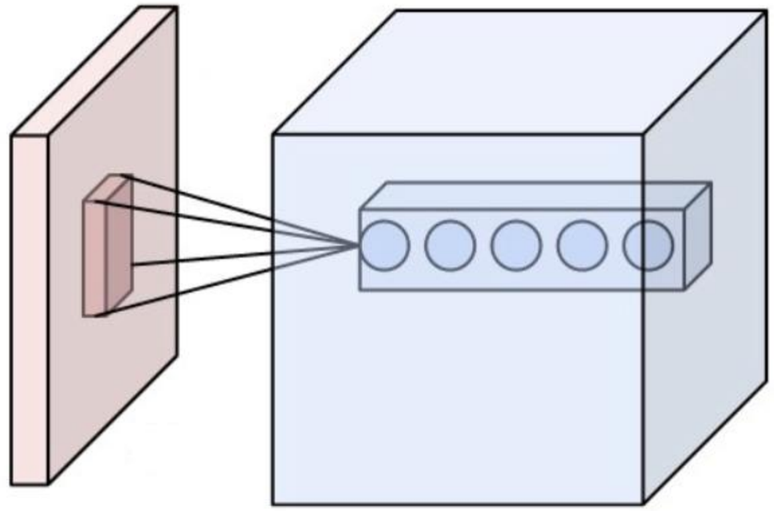


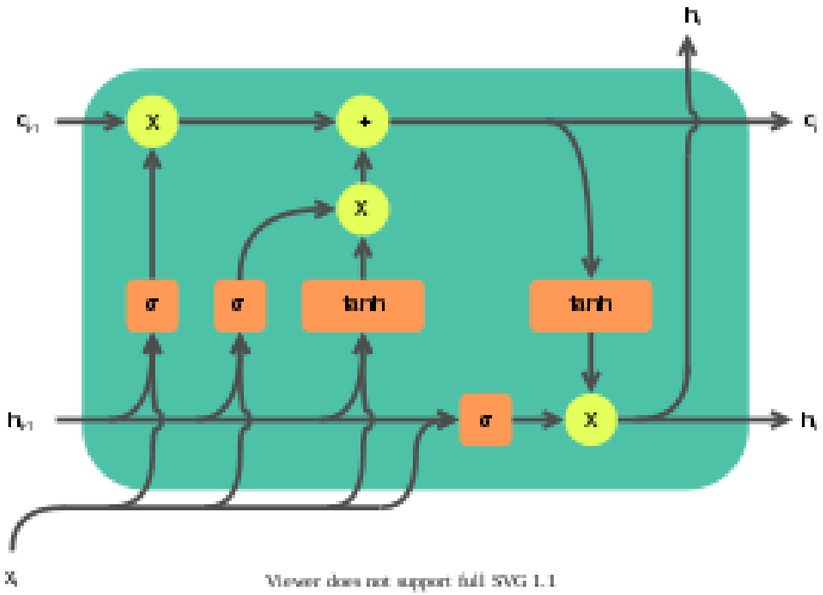


KI-Campus

Main Approaches in NLP

- Rule based approaches
  - Lack of enough accuracy
- Classical machine learning
  - Training data
  - Feature engineering
  - Training a model
- Deep learning
  - More training data
  - Feature engineering is skipped
  - Training a model





KI-Campus

Natural Language Processing

- Introduction to Natural Language Processing
- Everyday NLP applications
- Main NLP tasks
- Main approaches in NLP
- Who is this course for
- The Course structure
- NLP terminology

Who is this course for?

- Those who
  - don’t want use NLP models as a black box
    - Review the state-of-the-art approaches
  - want to gain intuition for problem solving
    - you’re asked to develop your own translation tool, what is the best approach
  - have a prior background on machine learning and deep learning

Supervised machine learning

Hidden layer

Backpropagation

softmax

LSTM

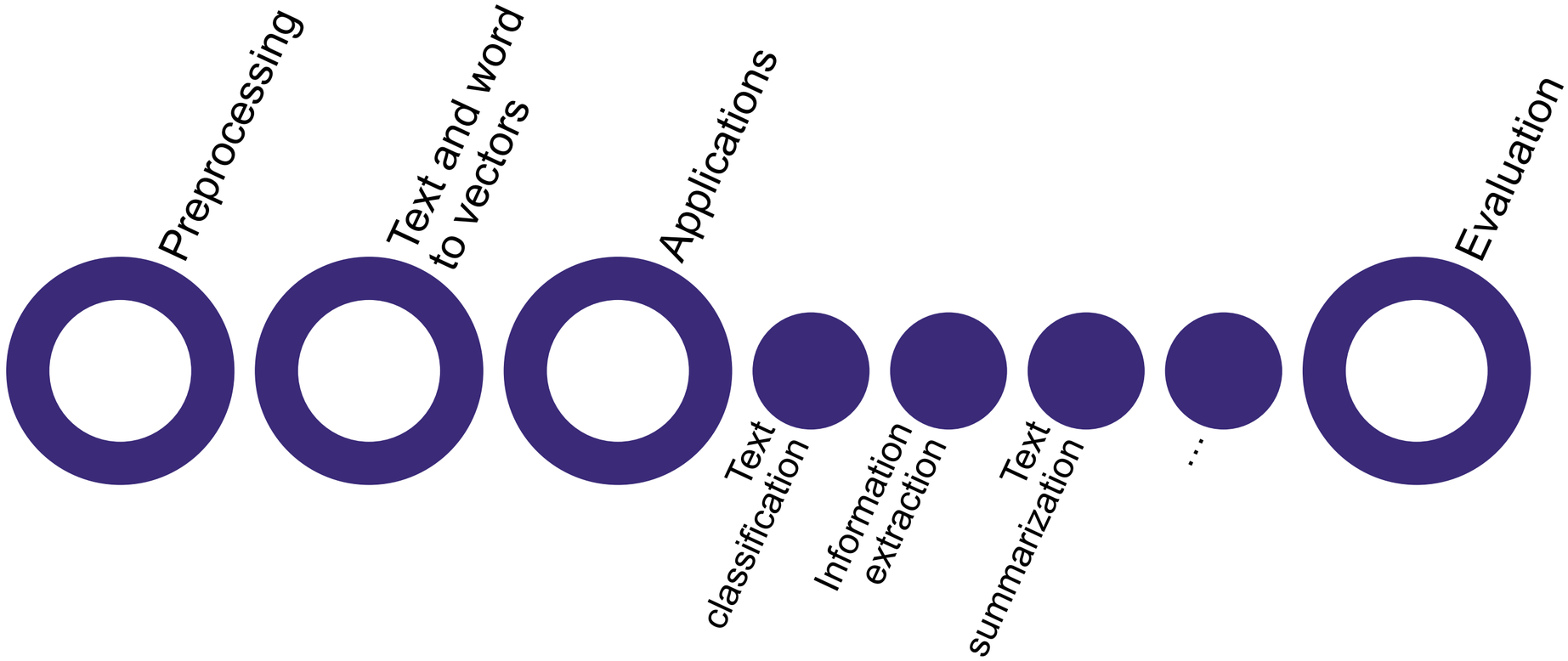
KI-Campus

Natural Language Processing

- Introduction to Natural Language Processing
- Everyday NLP applications
- Main NLP tasks
- Main approaches in NLP
- Who is this course for
- The Course structure
- NLP terminology

KI-Campus

The Course structure



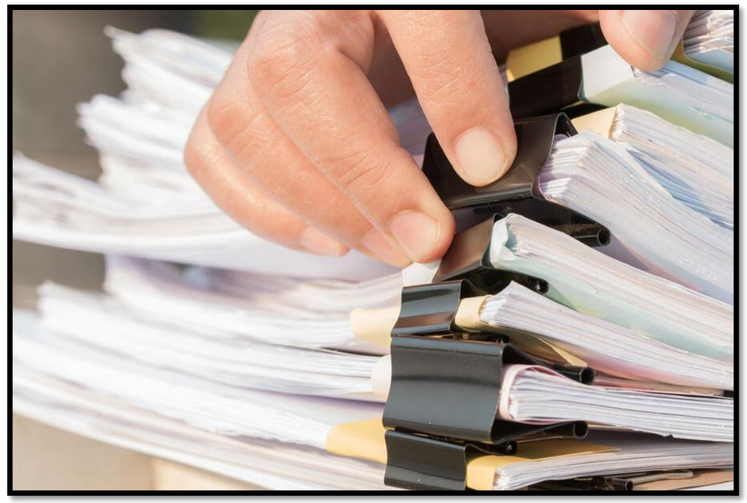
KI-Campus

Natural Language Processing

- Introduction to Natural Language Processing
- Everyday NLP applications
- Main NLP tasks
- Main approaches in NLP
- Who is this course for
- The Course structure
- NLP terminology

NLP terminology

- Corpus (Plural: corpora)
  - A collection of text, usually contains several documents
    - Wikipedia articles
    - Collection of movies reviews
    - Internet comments
    - Collection of tweets
  - Corpora can be in a single language or multiple languages



NLP terminology

- Document
  - Document refers to a body of text in a corpus
    - A tweet in a twitter corpus
    - An email in a collection of emails

- Stop word
  - usually refers to the most common words in a language
    - Words like ***“the”***, ***“and”***, ***“a”***, ***“an”***, ***“in”***

NLP terminology

- Vocabulary
  - The set of unique words used in the text corpus
    - Set of unique words which are used in all Wikipedia articles

- Out of Vocabulary (OOV)
  - Words that have not seen during the train, but in the test
  - We will encounter out of vocabulary terms when using our model for inference

KI-Campus

Thank you!

„KI-Campus – Die Lernplattform für Künstliche Intelligenz“ ist ein Projekt von

STIFTERVERBAND

HPI Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

NEOCOSMO

mmb Institut

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

www.ki-campus.org