
NLP for Software Engineering

Seminar im Sommersemester 2020
Vorbereitung, 28.01.2020



<http://www.broy.in.tum.de/lehre/seminare/SS20/NLP/>

Agenda

Zielsetzung des Seminars

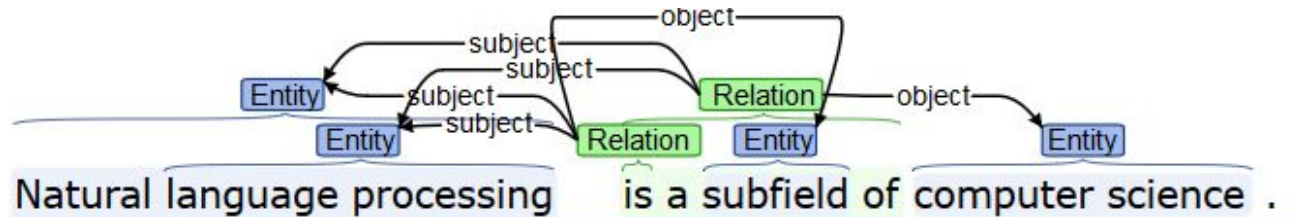
Ablauf des Seminars

Themen

Anmeldung

Zielsetzung des Seminars

- Anwendungsfelder von NLP ansehen
- Fokus auf Anwendung, weniger auf Grundlagen
- Konkrete Aufgabenstellungen
- selbst coden





Seminar Modus

9 Teilnehmer

Jeder Teilnehmer <-> ein Thema

- Selbstständiges Einarbeiten in Thema
- Diskussionen/Hilfestellung durch Betreuer
- Anwendung des Themas anhand Praxisbeispiel

Blockveranstaltung “Konferenz”

- Vorträge
- Diskussion



Voraussetzungen zur Teilnahme

- Grundlagen in NLP / Machine Learning / Deep Learning
 - *Natural Language Processing IN2361*
 - Eigene Projekte
 - ...
- Gute (mindestens passive) Englischkenntnisse
 - Literatur ist i.d.R. auf Englisch

Registration is open for TensorFlow Dev Summit 2020 [Learn more](#)

TensorFlow > Learn > TensorFlow Core > Tutorials ☆☆☆☆

Neural machine translation with attention

[Run in Google Colab](#)
[View source on GitHub](#)
[Download notebook](#)

This notebook trains a sequence to sequence (seq2seq) model for Spanish to English translation. This is an advanced example that assumes some knowledge of sequence to sequence models.

After training the model in this notebook, you will be able to input a Spanish sentence, such as "¿todavía estan en casa?", and return the English translation: "are you still at home?"

The translation quality is reasonable for a toy example, but the generated attention plot is perhaps more interesting. This shows which parts of the input sentence has the model's attention while translating:

★ **Note:** This example takes approximately 10 minutes to run on a single P100 GPU.

```

from __future__ import absolute_import, division, print_function, unicode_literals

import tensorflow as tf

import matplotlib.pyplot as plt
import matplotlib.ticker as ticker
from sklearn.model_selection import train_test_split

import unicodedata
import re
import numpy as np
import os
import io
import time

```

Download and prepare the dataset

We'll use a language dataset provided by <http://www.manythings.org/ankl/>. This dataset contains language translation

Bewertung

Vortrag (50%)

- Problemstellung
- Grundlagen
- Praxisteil
 - Ergebnisse, Erfahrungen, Probleme & Grenzen

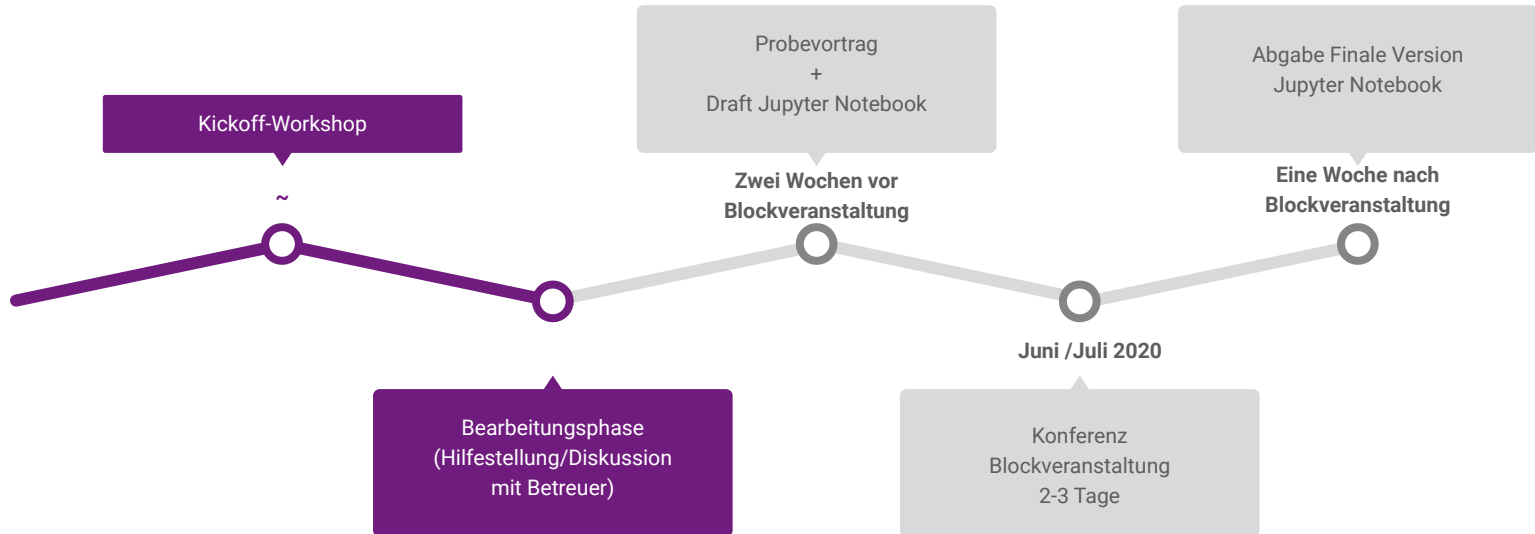
Jupyter-Notebook (50%)

- Grundlagen
- Erklärung des Praxisteils
- Code

Anwesenheitspflicht bei Konferenz



Ablauf Seminar

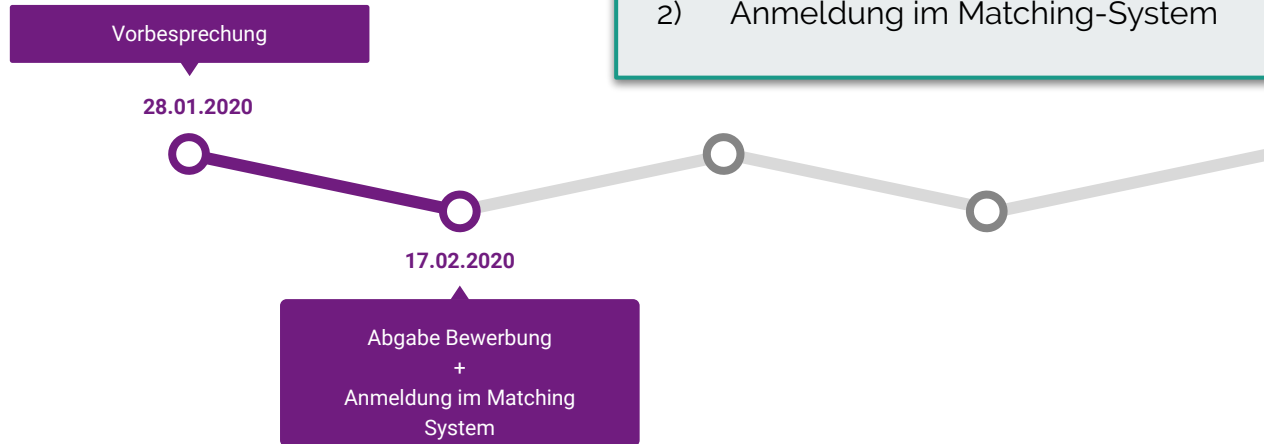




Anmeldung



Ablauf Anmeldung



Bewerbung:

1) Mail an junkerm@in.tum.de bis **17.2.2020**

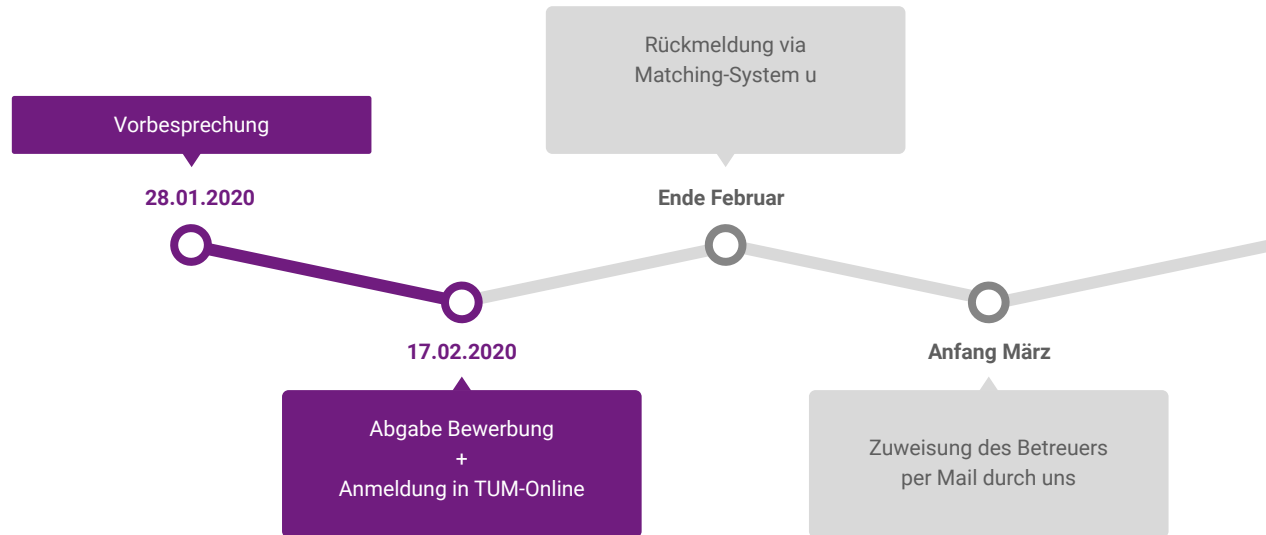
Inhalt

- Motivation für Seminar u. Themen (<1 Seite)
- Stand im Studium
- Erfahrungen (Arbeit, Praktikum, Open-Source-Projekt, ...)
- **Themenwünsche** (sortiert nach Priorität)

2) Anmeldung im Matching-System



Ablauf Anmeldung





Themen

Textklassifikation & Extraktion

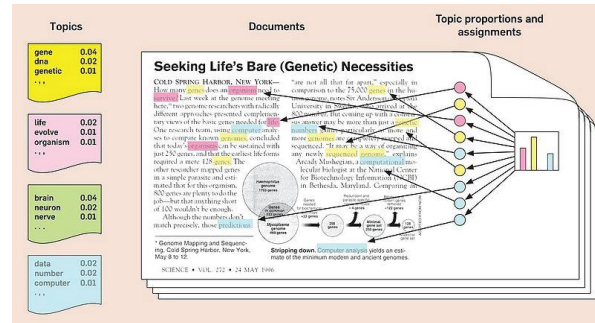
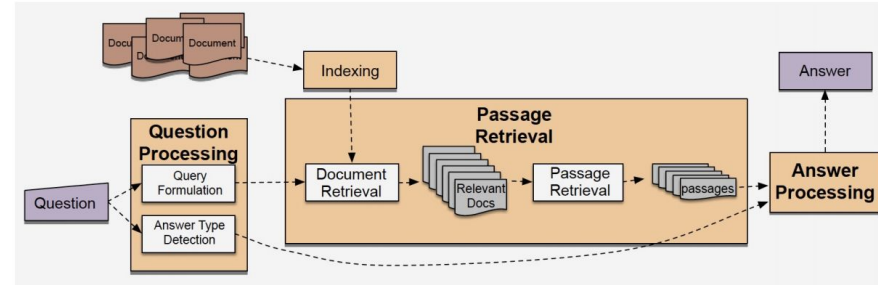
Thema 1: Sentiment-Analyse (Objectivity Analyse)

Thema 2: Question Answering / Information Extraction

Thema 3: Topic Extraction

Stichwörter:

- Binary-Classifier, IBM Watson, Bi-LSTMs, LDA, ...



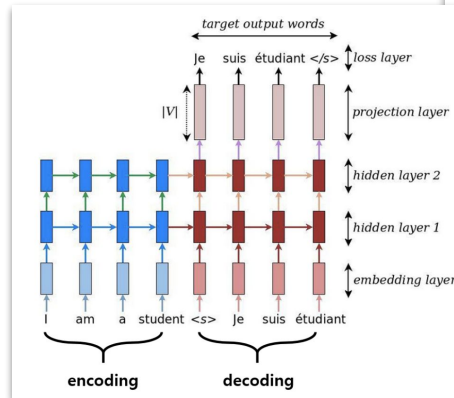
Text Transformation

Thema 4: Maschinelle Übersetzung

Thema 5: Automatische Textzusammenfassung

Stichwörter:

- Tensorflow, Encoder-Decoder, Attention, Transformer, seq2seq, ...



Talk to Transformer

See how a modern neural network completes your text. Type a custom snippet or try one of the examples. [Learn more](#) below.

[Follow @AdamDanielKing](#) for more neat neural networks.

Custom prompt

Natural language processing (NLP) is a subfield of linguistics, computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data.

GENERATE ANOTHER

Completion

Natural language processing (NLP) is a subfield of linguistics, computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data. Computers are capable of learning, processing, and speaking a wide variety of languages, including natural language processing. NLP is carried out in many different ways and techniques, including but not limited to programming languages, word algorithms, and natural language engines. The goals of NLP are to assist computers in the analysis and classification of natural language data, such as document data, text data, videos, images, and audio.

Data flow

Data flow is the sequence of operations that occur between a user's computer and

Detecting References & Relationships

Thema 6: Koreferenz Analyse

Thema 7: Extraktion von Objektreferenzen

Stichwörter:

- Dependency Parsing, Patterns, Neural Networks (BERT)

I voted for Obama because he was most aligned with my values", she said.

Rome was the capital of the Roman Empire and there were over 400 000 km of Roman Roads connecting the provinces to Rome. Augustus became the first Roman Emperor.



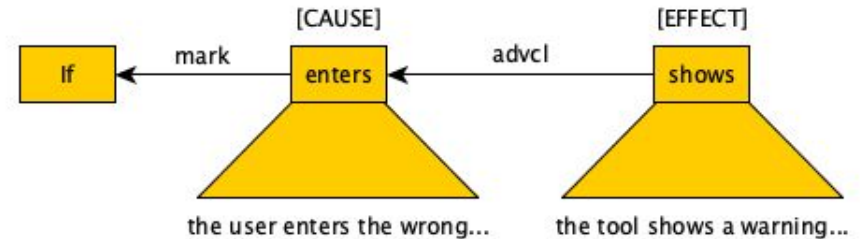
Anwendung auf Software Engineering

Thema 8: Erkennung von Ursachen und Wirkungen in Software Anforderungen

Thema 9: Suchen von semantischen Duplikaten in Software Anforderungen

Stichwörter:

- Dependency Parsing, Extraction Rules, Neural Networks
- Word/Sentence Embeddings



Fragen?



<http://www.broy.in.tum.de/lehre/seminare/SS20/NLP/>