

Natural Language Processing II

Course code: MLC_NLPA

In this course, we will follow up the basic course on Natural Language Processing with advanced topics. We will mainly focus on text data preprocessing and state-of-the-art applications of deep learning in NLP. We will particularly work with so-called Transformers. By using the transfer learning technique we will show how to exploit large pre-trained neural networks for various practical applications.

Required skills

- basic knowledge of programming in Python
- high school level of mathematics
- Basics of machine learning on the level of our course Introduction to machine Learning
- Knowledge on the level of our basic Natural Language Processing course

Course outline

- Preprocessing
- a few words about encoding, unicode normalization
- traditional tokenization (simple methods, spacy, moses)
- subword tokenization (byte-pair encoding, wordpiece, sentencepiece)
- cleaning (deduplication, boilerplate removal)
- Word embeddings
- universal ideas
- skip-gram implementation
- Machine Translation with RNN
- LSTM and GRU overview
- RNN language translation implementation
- Transformers
- attention is all you need
- transformer architecture
- GPT2
- BERT
- XLNET
- Examples of transfer learning in NLP
- text classification
- named entity recognition
- question answering

GOPAS Praha

Kodaňská 1441/46
101 00 Praha 10
Tel.: +420 234 064 900-3
info@gopas.cz

GOPAS Brno

Nové sady 996/25
602 00 Brno
Tel.: +420 542 422 111
info@gopas.cz

GOPAS Bratislava

Dr. Vladimíra Clementisa 10
Bratislava, 821 02
Tel.: +421 248 282 701-2
info@gopas.sk



Copyright © 2020 GOPAS, a.s.,
All rights reserved