Zpracování přirozeného jazyka II

Kód kurzu: MLC_NLPA

V tomto kurzu navážeme na základní kurz Zpracování přirozeného jazyka pokročilejšími tématy. Zaměříme se především na předzpracování dat a nejnovější aplikace hlubokého učení ve zpracování textu. Bude se jednat především architektury neuronových sítí postavených na takzvaných Transformerech. S využitím metody transfer learningu ukážeme, jak lze využít velké předtrénované neuronové sítě pro nejrůznější praktické aplikace.

Požadované vstupní znalosti

- Základní znalost programování v Pythonu
- Středoškolská matematika
- Znalosti strojového učení na úrovni kurzu Úvod do strojového učení
- Znalosti na úrovni kurzu Zpracování přirozeného jazyka

Studijní materiály

Studijní materiál společnosti Machine Learning College.

Osnova kurzu

- Předzpracování textových dat
- Kódování znaků a unicode normalizace
- Tradiční tokenizace (jednoduché metody, Spacy, Moses)
- Subword tokenizace (byte-pair kódování, wordpiece, sentencepiece)
- Šištění dat (deduplikace, odstranění textového balastu)
- Word embeddings
- Obecné principy
- Implementace skip-gram modelu
- Strojový překlad s rekurentními sítěmi
- Paměťové buňky LSTM a GRU
- Implementace strojového překladu pomocí rekurentních sítí
- Transformery
- Attention is all you need
- Architektura transformeru
- GPT2
- BERT
- XLNET
- Příklady transfer learningu pro zpracování přirozeného jazyka
- Klasifikace textů
- Rozpoznání jmenných entit
- Question answering

GOPAS Praha
Na Strži 2097/63
140 00 Praha 4 - Krč
Tel.: +420 226 201 390
info@gopas.cz

GOPAS Brno
Nové sady 996/25
602 00 Brno
Tel.: +420 530 513 590
info@gopas.cz

GOPAS Bratislava
Dr. Vladimíra Clementisa 10
Bratislava, 821 02
Tel.: +421 902 903 132
info@gopas.sk

Copyright © 2026 GOPAS, a.s.,
All rights reserved