

# Zpracovanie prirodzeného jazyka II

Kód kurzu: MLC\_NLPA

V tomto kurze nadviažeme na základný kurz Spracovanie prirodzeného jazyka pokročilejšími témami. Zameriame sa hlavne na predspracovanie dát a najnovšie aplikácie hlbokého učenia v spracovaní textu. Bude sa jednať predovšetkým o architektúry neurónových sietí postavených na tzv. Transformeroch. S využitím metódy transfer learningu ukážeme, ako možno využiť veľké predtrénované neurónové siete pre najrôznejšie praktické aplikácie.

## Požadované vstupné znalosti

- Základná znalosť programovania v Pythone
- Stredoškolská matematika
- Znalosti strojového učenia na úrovni kurzu Úvod do strojového učenia
- Znalosti na úrovni kurzu Spracovanie prirodzeného jazyka

## Študijné materiály

Študijný materiál spoločnosti Machine Learning College.

## Osnova kurzu

- Predspracovanie textových dát
- Kódovanie znakov a normalizácia Unicode
- Tradičná tokenizácia (jednoduché metódy, Spacy, Moses)
- Subword tokenizácia (byte-pair kódovanie, wordpiece, sentencepiece)
- Čistenie dát (duplikácia, odstránenie textového balastu)
- Word embeddings
- Všeobecné princípy
- Implementácia skip-gram modelu
- Strojový preklad s rekurentnými sieťami
- Pamäťové bunky LSTM a GRU
- Implementácia strojového prekladu pomocou rekurentných sietí
- Transformer
- Attention is all you need
- Architektúra transformera
- GPT2
- BERT
- XLNET
- Príklady transfer learningu pre spracovanie prirodzeného jazyka
- Klasifikácia textov
- Rozpoznanie menovitých entít
- Question answering

### GOPAS Praha

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

### GOPAS Brno

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

### GOPAS Bratislava

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



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