

# Zpracování přirozeného jazyka

Kód kurzu: MLC\_NLP

Kurz je zaměřen na analýzu a zpracování textů. Předpokládá se znalost principů strojového učení, ale ty nejdůležitější koncepty budou stručně zopakovány. Specifikem zpracování textů je způsob předzpracování dat a jejich vektorizace. Tomu bude věnována první část. Vše bude prakticky vyzkoušeno na úloze, jejíž cílem je klasifikace textových dokumentů. Dále se účastníci dozvědí, co jsou to jazykové modely a jak je použít pro detekci jazyka dokumentu nebo generování textů.

| Pobočka    | Dnů | Cena kurzu | ITB |
|------------|-----|------------|-----|
| Praha      | 1   | 4 990 Kč   | 0   |
| Bratislava | 1   | 210 €      | 0   |

Uvedené ceny jsou bez DPH.

## Termíny kurzu

| Datum | Dnů | Cena kurzu | Typ výuky | Jazyk výuky | Lokalita |
|-------|-----|------------|-----------|-------------|----------|
|-------|-----|------------|-----------|-------------|----------|

Uvedené ceny jsou bez DPH.

## Požadované vstupní znalosti

- Základní znalost programování v Pythonu
- Středoškolské znalosti lineární algebry, matematické analýzy a teorie pravděpodobnosti. Bude předpokládáno základní porozumění pojmům jako vektor, matice, vektorový prostor, pravděpodobnost, podmíněná pravděpodobnost, nezávislost náhodných jevů a znalost násobení matic a derivace funkcí.
- Znalosti strojového učení na úrovni kurzu Úvod do strojového učení.

## Studijní materiály

Studijní materiál společnosti Machine Learning College.

## Osnova kurzu

- Úvod do zpracování přirozeného jazyka
- Vybrané kapitoly z počítačové lingvistiky (korpuse, tokenizace, morfologická, syntaktická a sémantická analýza, entropie, mutual information, perplexita)
- Vektorizace textových dokumentů (bag of words, one-hot encoding, TF-IDF)
- Word embedding (word2vec)
- Praktická úloha na klasifikaci textů
- Word embedding (vytvoření word2vec modelů a experimenty s vektorovými reprezentacemi slov)
- Úvod do jazykových modelů (n-gramové modely, vyhlazování, modely založené na neuronových sítích)
- Praktická úloha na jazykové modelování (implementace jazykových modelů a jejich využití pro detekci jazyka textu)
- Úprava algoritmu pro generování textů

**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