

## KIV/UIR - Semestrální práce pro ak. rok 2018/19

### Automatická klasifikace pojmenovaných entit

Ve zvoleném programovacím jazyce navrhnete a implementujete program, který umožní automaticky určit, zda je ve větě pojmenovaná entita, či nikoli. V případě, že je, určete její typ. Při řešení budou splněny následující podmínky:

- Pojmenované entity jsou hlavní významová slova a slovní spojení ve větě jako např. jména osob, institucí, geografické názvy, časové údaje, atp.
- Použijte data z českého historického periodika „Posel od Čerchova“, která jsou k dispozici na [https://drive.google.com/drive/u/1/folders/14ZX-r5-R\\_Fud\\_x57LZf0m5qsbpn8QMBN](https://drive.google.com/drive/u/1/folders/14ZX-r5-R_Fud_x57LZf0m5qsbpn8QMBN). V původní podobě jsou data k dispozici na <http://www.portafontium.eu/periodical/posel-od-cerchova-1872?language=cs>.
- Pro vyhodnocení přesnosti implementovaných algoritmů bude NUTNÉ pojmenované entity ručně označovat. Každý student ručně anotuje cca 3 strany zadaného textu – *termín 31.3.2019*.
- Přiřazení konkrétního textu jednotlivým studentům spolu s návodem na anotaci a příklady je uloženo spolu s daty na výše uvedené adrese.
- Implementujte alespoň dva různé algoritmy (z přednášek i vlastní) pro tvorbu příznaků reprezentující text.
- Implementujte alespoň dvě různé metody klasifikace dle vlastní volby s využitím metod reprezentace textu viz výše.
- Funkčnost programu bude následující:
  - Spuštění s parametry: množina\_obsahující\_pojmenované\_entity, parametrizační\_algoritmus, algoritmus\_klasifikace, (další\_volitelné\_parametry)
  - Program provede klasifikaci pojmenovaných entit dle zadaných parametrů, výsledky klasifikace uloží do souboru a zároveň vyhodnotí úspěšnost klasifikace.
- Vyhodnoťte kvalitu klasifikačního algoritmu na anotovaných datech. Použijte metriky přesnost (accuracy). Otestujte všechny konfigurace programu (tedy celkem 4 výsledky).
- Bonusový úkol:
  - Implementujte klasifikační algoritmus pomocí neuronové sítě typu LSTM (Long Short-Term Memory) s využitím knihoven Keras a Tensorflow.