

Propozycja tematu pracy dla doktoranta:

Wnioskowanie i weryfikacja informacji tekstowych za pomocą metod uczenia maszynowego.

dr hab. inż. Szymon Jaroszewicz
Instytut Podstaw Informatyki PAN
s.jaroszewicz@ipipan.waw.pl

Opis projektu:

Wnioskowanie z języka naturalnego (ang. natural language inference) jest obecnie jedną z najintensywniej rozwijających się i jednocześnie jedną z najtrudniejszych dziedzin przetwarzania języka naturalnego. Polega na wyciąganiu wniosków z tekstu i weryfikacji hipotez w oparciu o nie, co może być nieocenioną pomocą przy np. sprawdzaniu wiarygodności wiadomości medialnych lub pojawiających się w mediach społecznościowych. Obecnie do tego zadania używa się głównie uczenia maszynowego z nadzorem - ogromnych modeli wytrenowanych na wielkich ilościach danych tekstowych. Największą wadą tego typu modeli jest brak implementacji logiki, co powoduje, że zachowanie tych metod w przypadku nowych, niepodobnych do wcześniejszych danych jest nieprzewidywalne. Celem projektu jest, korzystając z obecnych metod, zbadać i zaimplementować modele, które używają logiki przy wnioskowaniu z tekstu. Rezultatem projektu, poza rozprawą doktorską, będzie biblioteka, zawierająca implementację nowych metod. Ponadto planowane jest przeprowadzenie eksperymentów i, celem porównania proponowanych rozwiązań z innymi konkurencyjnymi systemami, wzięcie udziału w konkursach detekcji fake-newsów.

Oczekiwania:

1. Wykształcenie wyższe: informatyka lub pokrewne.
2. Dobra znajomość uczenia maszynowego w przetwarzaniu języka naturalnego
3. Programowanie: znajomość Pythona, dodatkowym atutem będzie znajomość języków C++ oraz doświadczenie z systemami Linux
4. Dobra znajomość języka angielskiego

Literatura:

1. Martin, James H., and Daniel Jurafsky. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River: Pearson/Prentice Hall, 2009.
2. Zhou, Xinyi, and Reza Zafarani. "Fake news: A survey of research, detection methods, and opportunities." *arXiv preprint arXiv:1812.00315* (2018).
3. Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).