

Inteligentne przetwarzanie języka naturalnego

Rafał Jaworski, specjalność Sztuczna Inteligencja

1. Charakterystyka problemu badawczego

Prowadzone przeze mnie badania koncentrują się wokół algorytmów przetwarzania języka naturalnego. Jest to zbiór algorytmów i technik umożliwiających między innymi analizę, modelowanie, rozumienie oraz tłumaczenie tekstów w naturalnych językach ludzkich, na przykład polskim lub angielskim. Obszar badawczy obejmuje usprawnianie oraz opracowywanie nowych algorytmów zdolnych analizować i przetwarzać tekst, jak również rozszerzenie istniejących algorytmów pod kątem wsparcia nowych języków.

2. Motywacja/opis instytucji, proponującej współpracę nad projektem

Motywację do wyżej wymienionych badań stanowi obserwowany wzrost zainteresowania naukowego technikami przetwarzania języka naturalnego, w szczególności w zakresie wielkich modeli językowych (Large Language Models). Możliwości i potencjalne zastosowanie tych technik nie zostało jeszcze w pełni zbadane. Magistrant podejmujący się wykonania projektu badawczo-rozwojowego oraz przygotowania pracy magisterskiej powinien zaprojektować autorskie rozwiązanie wybranego problemu oraz wykonać serię eksperymentów. Celem tych eksperymentów winna być ewaluacja istniejących algorytmów oraz odniesienie ich do rozwiązania autorskiego.

3. Tematy i opis projektów

- programy tłumaczeniowe - wspomaganie pracy tłumacza poprzez korekcję tekstu, zarządzanie terminologią, optymalne wykorzystanie pamięci tłumaczeniowej
- analiza wydźwięku tekstu - analiza i automatyczna ocena postów w mediach społecznościowych, opinii o produktach
- badania lingwistyczne - wyszukiwanie i automatyczna anotacja wybranych form słownych w tekstach w wielu językach
- klasyfikacja tekstu na podstawie wybranych cech
- wykorzystanie Large Language Models

4. Wymagania odnośnie członków projektu

Od studentów podejmujących się wykonania pracy magisterskiej w wyżej opisanym zakresie wymaga się biegłości w programowaniu w dowolnie wybranym obiektowym języku programowania i znajomości efektywnych algorytmów, znajomości technik uczenia maszynowego oraz znajomości podstawowych technik przetwarzania języka naturalnego: tokenizacja, stemming, POS-tagging.

5. Literatura

- [1] Géron, Aurélien. 2019. Hands-on Machine Learning with Scikit-Learn, Keras and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. 2nd ed. CA 95472: O'Reilly.
- [2] Indurkha, Nitin and Damerau, Fred J.. Handbook of Natural Language Processing. 2nd : Chapman & Hall/CRC, 2010.
- [3] Aggarwal, Charu C.. Neural Networks and Deep Learning. Cham: Springer, 2018.