

Technologie programistyczne

Dominik Doberski 133207, Michał Kukieła 132265

Temat projektu

Podział tekstu na klasy (opis przyrody i pozostałe) za pomocą uczenia maszynowego.

Cel

Stworzenie klasyfikatora zdolnego podzielić tekst na części będące opisem przyrody i pozostałe.

Umożliwi to na przykład skrócenie szkolnych lektur o zbędne (z technicznego punktu widzenia, pomijając aspekty literackie), nieraz długie i rozbudowane opisy przyrody, przy zachowaniu fabuły i istotnych treści tychże lektur.

Języki i narzędzia

Język: Python.

Narzędzia: Biblioteki do uczenia maszynowego (tensorflow)

Realizacja

1. Stworzenie zbioru uczącego i testującego
2. Tekst spróbujemy klasyfikować za pomocą naiwnego klasyfikatora bayesowskiego.
3. Jeżeli takie rozwiązanie nie będzie wystarczająco skuteczne lub zostanie dużo czasu spróbujemy stworzyć sieć neuronową klasyfikującą tekst, lub ewentualnie poszukać jeszcze innych metod.