  
  
**Eksploracja danych tekstowych - laboratoria 8: Rozpoznawanie encji nazwanych NER**

*Do realizacji zadań wykorzystaj środowisko jupyter notebook. Po zakończeniu zadań wyślij odpowiedzi na adres michal.zarnecki@cdv.pl.*

**1 Rozpoznawanie jednostek nazwanych**

Manualnie wyodrębnij z poniższego tekstu następujące jednostki nazwane: organizacja, data, czas, osoba, lokalizacja (np. miasta, kraje) oraz lokalizacja geograficzna (jak: rzeki, góry itp.).

*"31 grudnia każdego roku o godzinie 11:00 czasu polskiego, jako pierwsi świętują przyjście Nowego Roku mieszkańcy Republiki Kiribati.*

*Republika Kiribati to państwo wyspiarskie położone na obszarze Oceanii, złożone z 33 atoli rozsianych po Pacyfiku. Najważniejszą wyspą w archipelagu jest South Tarawa, pełniąca rolę stolicy i zamieszkiwana przez większość ludności. Nazwa "Kiribati" jest miejscową wersją słowa "Gilberts" pochodzącego z języka angielskiego, które odnosi się do Wysp Gilberta, jednej z głównych grup wysp w tym kraju.*

*Jednym ze wybitnych obywateli Kiribati jest Anote Tong, były prezydent kraju. Zyskał on międzynarodową sławę dzięki swojemu zaangażowaniu w działania związane ze zmianami klimatu. Innym znanym przywódcą Kiribati był Teburoro Tito, były prezydent, który przyczynił się do promocji i obrony praw Kiribati na arenie międzynarodowej.*

*Na Kiribati istnieje wiele organizacji, które odgrywają kluczową rolę w różnych aspektach życia społecznego i kulturalnego. Przykładem jest Ueen Te Makeke, która działa na rzecz promocji kultury i zachowania tradycji Kiribati. Red Cross Society dostarcza pomoc humanitarną i medyczną mieszkańcom Republiki, szczególnie ważną w obliczu zagrożeń związanych ze zmianami klimatycznymi. Natomiast Te Waa Mai skupia się na ochronie środowiska naturalnego oraz zrównoważonym rozwoju archipelagu.*

*Niestety, ze względu na globalne zmiany klimatyczne i podnoszenie się poziomu morza, Kiribati zmaga się z problemem migracji ludności, a w przyszłości może całkowicie zniknąć pod powierzchnią wody."*

**3. Skorzystaj z biblioteki spacy aby sprawdzić jak model “pl\_core\_news\_lg” wykryje** jednostki nazwane w powyższym tekście.

Czy wykryłeś encje tak samo jak model? Jak myślisz, w jaki sposób można polepszyć efektywność modelu wykrywania jednostek nazwanych?

**4. Parsowanie zależnościowe**

1. Przy wykorzystaniu jupyter notebooka, stwórz drzewo składniowe dla zdania: “Zdrowie to podstawa, zacznij od skóry”.

2. Jaka jest relacja pomiędzy słowami “podstawa” i “zdrowie”? Tutaj znajdziesz objaśnienie pojęć relacji składniowych w języku angielskim: https://universaldependencies.org/en/dep/