

#### ZADANIE 5: Prosty Web Scrapper nagłówków różnych stron WWW:

1. Utworzyć wirtualne środowisko, w którym będą instalowane zewnętrzne moduły pythona.
2. Z wykorzystaniem modułów do Web Scrappingu, stworzyć kod w Pythonie do pobierania tytułów/nagłówków artykułów z różnych stron internetowych (min. 25 stron) - zastosować wielowątkowość i obsługę wyjątków. W celu sprawdzenia poprawności danych, wyświetlić przykładowe nagłówki w konsoli. Strony należy pogrupować według rodzaju, np. strony informacyjne, sportowe, plotkarskie, itd. Jedna strona WWW może znajdować się w kilku grupach (np. informacyjna i sportowa).
3. Dodać funkcjonalność, dzięki której program będzie zliczał słowa w tytułach/nagłówkach dla różnych kategorii serwisów - można zastosować klasę Counter z modułu collections
4. Stworzyć wizualizację graficzną najczęściej występujących słów w nagłówkach stron dla poszczególnych kategorii serwisów - wykresy słupkowe, kołowe, itd.
5. Utworzyć dokumentację kodu (tylko docstringi)

Kryteria oceniania:

dst – pobieranie nagłówków bez wielowątkowości, wyświetlenie nagłówków w konsoli

db – pobieranie nagłówków z wykorzystaniem wielowątkowości, zliczanie słów dla różnych kategorii serwisów WWW

bdb – powyższe + wizualizacja graficzna wyników

Zadanie dla chętnych (na kolejną 5) - stworzenie modelu (np. z wykorzystaniem algorytmów uczenia maszynowego) do klasyfikacji stron WWW (przypisanie strony z poza listy „uczącej” model do konkretnej kategorii, np. sportowa, informacyjna)