

## Laboratorium 2

Proszę przesłać rozwiązania do Portalu Edukacyjnego w pliku z nazwą: nazwiskoLab2.py

Zadania.

1. Dla text5 z NLTK book sprawdź ile jest różnych bigramów. Dla 10 najczęściej występujących podaj ich liczbę wystąpień i porównaj z częstością 10 najczęściej występujących słów.
2. Dla text5 wyznacz liczbę n-gramów dla  $n=3,4,5,6$ .
3. Dla wyników z poprzedniego zadania sporządź wykres.
4. Proszę przetestować działanie pakietu re do obsługi wyrażeń regularnych.  
Znalezc:
  - (a) wszystkie wystąpienia słów z końcówką 'ing' w text1,
  - (b) policzyć liczbę wystąpień samogłosek w text2,
  - (c) policzyć liczbę wystąpień skrótu U.S.A w text4.
5. Zdefiniuj funkcję, która dla tekstu zlicza jaki procent stanowią słowa spoza listy stopwords. Zastosuj funkcję do każdego z tekstów z NLTK Book.
6. Zdefiniować funkcję, która dla zadanego zdania wyszukuje 3-gramy pasujące do wzorca czasownik+TO+czasownik. Przetestować na 5 wybranych przykładach.