Laboratorium 1

Częśc I.
Rozpoczynamy od instalacji Pythona, PyCharm i biblioteki NLTK. Ewentualnie, mozna korzystać z Anaconda.
Część II.
Korzystamy z biblioteki NLTK. Jako pierwszy należy wykonć skrypt służący do załadowania biblioteki oraz danych z NLTK Book:

```
import nltk
nltk.download()
from nltk.book import *
```

Część III.
Poniżej przeanalizujemy przykłady ilustrujace działanie wybranych użytecznych funkcji:

- funkcja *len()* zwraca liczbę wystąpień tokenów, czyli ciągów znaków (słowa, znaki specjalne np. emotikony); złożenie *len(set())* zwraca liczbę wystąpień bez powtórzeń,
- funkcja *count()* zwraca liczbę wystąpień ustalonego słowa,
- metoda *similar()* z parametrem, będącym słowem wykonana na tekscie zwraca słowa podobne,
- metoda *concordance()* z parametrem, będacym słowem zwraca występowania słowa wraz z kontekstem,
- metoda *common_contexts()* dla listy słów zwraca konteksty występowania słow z listy

Przykłady do przetestowania:

```
print(len(text3))
print(len(set(text3)))
text1.count("monstrous")
text1.similar("monstrous")
text1.concordance("monstrous")
text1.common_contexts(["stories", "pictures"])
```

Czesć IV.
Zadania.

1. Dla każdego z tekstów z NLTK Book należy wybrać dwa słowa i napisać kod, który zwraca wspólne konteksty występowania tych słów.

2. Korzystając z funkcji lexical_diversity (len(text) / len(set(text))- liczba liter w tekscie podzielona przez liczbę różnych liter) uzupełnić poniższą tabelkę dla danych z nltk.book

tekst	liczba słów	słowa różne	lexical_diversity
text1	260819	19317	13.50
text2			
text3			
text4			
text5			

3. Stwórz listę wszystkich słów 4-literowych z text1. Ile ich jest?

4. W text1 znajdź wszystkie wystąpienia słów długości większej niż 17.

5. Korzystając z funkcji set i sorted wyznacz słownik dla każdego ze zdań sent1, . . . , sent8 oraz wspólny słownik dla wszystkich wymienionych zdań.

6. Podaj definicję funkcji VocabSize(), która dla tekstu zwraca rozmiar słownika - czyli wylicza ile jest słów różnych. Zastosuj do każdego z tekstów z nltk.book.

7. Wyznacz 10 najczęsciej występujących słów w text1.

8. Sprawdź jakie są najdłuższe słowa w każdym z tekstów text1,..., text6.

Rozwiązania część IV zad 1-8 proszę przesłać do Portalu Edukacyjnego w pliku z rozszerzeniem .py i nazwą:
nazwiskoLab1.py