Лабораторна робота #2

Попередня обробка тексту за допомогою NLTK

Мета роботи: Ознайомитись з основними операціями з попередньої обробки тексту та їх реалізацією у бібліотеці NLTK.

Варіант 1. ІП-13 Ал Хадам Мурат

1. Зчитати файл text1. а) Порахувати кількість речень в тексті; б) вивести 10 слів, які зустрічаються найчастіше; в) провести лематизацію слів третього речення (попередньо визначивши частини мови).
2. Використати корпус Brown, перший текст категорії fiction. а) Вивести перші 5 речень; б) Вивести 10 іменників, що зустрічаються найчастіше.

```
import nltk
nltk.download('all')
```

```
[nltk_data] Downloading collection 'all'
[nltk_data]    |
[nltk_data]    | Downloading package abc to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package abc is already up-to-date!
[nltk_data]    | Downloading package alpino to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package alpino is already up-to-date!
[nltk_data]    | Downloading package averaged_perceptron_tagger to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package averaged_perceptron_tagger is already up-
[nltk_data]    |       to-date!
[nltk_data]    | Downloading package averaged_perceptron_tagger_ru to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package averaged_perceptron_tagger_ru is already
[nltk_data]    |       up-to-date!
[nltk_data]    | Downloading package basque_grammars to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package basque_grammars is already up-to-date!
[nltk_data]    | Downloading package bcp47 to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package bcp47 is already up-to-date!
[nltk_data]    | Downloading package biocreative_ppi to
```

```
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package biocreative_ppi is already up-to-date!
[nltk_data]     | Downloading package bllip_wsj_no_aux to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package bllip_wsj_no_aux is already up-to-date!
[nltk_data]     | Downloading package book_grammars to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package book_grammars is already up-to-date!
[nltk_data]     | Downloading package brown to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package brown is already up-to-date!
[nltk_data]     | Downloading package brown_tei to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package brown_tei is already up-to-date!
[nltk_data]     | Downloading package cess_cat to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package cess_cat is already up-to-date!
[nltk_data]     | Downloading package cess_esp to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package cess_esp is already up-to-date!
[nltk_data]     | Downloading package chat80 to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package chat80 is already up-to-date!
[nltk_data]     | Downloading package city_database to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package city_database is already up-to-date!
[nltk_data]     | Downloading package cmudict to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package cmudict is already up-to-date!
[nltk_data]     | Downloading package comparative_sentences to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package comparative_sentences is already up-to-
[nltk_data]     |       date!
[nltk_data]     | Downloading package comtrans to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package comtrans is already up-to-date!
[nltk_data]     | Downloading package conll2000 to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package conll2000 is already up-to-date!
[nltk_data]     | Downloading package conll2002 to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package conll2002 is already up-to-date!
[nltk_data]     | Downloading package conll2007 to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package conll2007 is already up-to-date!
[nltk_data]     | Downloading package crubadan to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package crubadan is already up-to-date!
[nltk_data]     | Downloading package dependency_treebank to
```

```
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package dependency_treebank is already up-to-date!
[nltk_data]     | Downloading package dolch to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package dolch is already up-to-date!
[nltk_data]     | Downloading package europarl_raw to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package europarl_raw is already up-to-date!
[nltk_data]     | Downloading package extended_omw to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package extended_omw is already up-to-date!
[nltk_data]     | Downloading package floresta to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package floresta is already up-to-date!
[nltk_data]     | Downloading package framenet_v15 to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package framenet_v15 is already up-to-date!
[nltk_data]     | Downloading package framenet_v17 to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package framenet_v17 is already up-to-date!
[nltk_data]     | Downloading package gazetteers to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package gazetteers is already up-to-date!
[nltk_data]     | Downloading package genesis to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package genesis is already up-to-date!
[nltk_data]     | Downloading package gutenberg to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package gutenberg is already up-to-date!
[nltk_data]     | Downloading package ieer to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package ieer is already up-to-date!
[nltk_data]     | Downloading package inaugural to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package inaugural is already up-to-date!
[nltk_data]     | Downloading package indian to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package indian is already up-to-date!
[nltk_data]     | Downloading package jeita to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package jeita is already up-to-date!
[nltk_data]     | Downloading package kimmo to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package kimmo is already up-to-date!
[nltk_data]     | Downloading package knbc to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package knbc is already up-to-date!
[nltk_data]     | Downloading package large_grammars to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
```

```
[nltk_data]     |   Package large_grammars is already up-to-date!
[nltk_data]     | Downloading package lin_thesaurus to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package lin_thesaurus is already up-to-date!
[nltk_data]     | Downloading package mac_morpho to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package mac_morpho is already up-to-date!
[nltk_data]     | Downloading package machado to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package machado is already up-to-date!
[nltk_data]     | Downloading package masc_tagged to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package masc_tagged is already up-to-date!
[nltk_data]     | Downloading package maxent_ne_chunker to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package maxent_ne_chunker is already up-to-date!
[nltk_data]     | Downloading package maxent_treebank_pos_tagger to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package maxent_treebank_pos_tagger is already up-
[nltk_data]     |       to-date!
[nltk_data]     | Downloading package moses_sample to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package moses_sample is already up-to-date!
[nltk_data]     | Downloading package movie_reviews to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package movie_reviews is already up-to-date!
[nltk_data]     | Downloading package mte_teip5 to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package mte_teip5 is already up-to-date!
[nltk_data]     | Downloading package mwa_ppdb to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package mwa_ppdb is already up-to-date!
[nltk_data]     | Downloading package names to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package names is already up-to-date!
[nltk_data]     | Downloading package nombank.1.0 to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package nombank.1.0 is already up-to-date!
[nltk_data]     | Downloading package nonbreaking_prefixes to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package nonbreaking_prefixes is already up-to-date!
[nltk_data]     | Downloading package nps_chat to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package nps_chat is already up-to-date!
[nltk_data]     | Downloading package omw to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package omw is already up-to-date!
[nltk_data]     | Downloading package omw-1.4 to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
```

```
[nltk_data]     |   Package omw-1.4 is already up-to-date!
[nltk_data]     | Downloading package opinion_lexicon to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package opinion_lexicon is already up-to-date!
[nltk_data]     | Downloading package panlex_swadesh to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package panlex_swadesh is already up-to-date!
[nltk_data]     | Downloading package paradigms to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package paradigms is already up-to-date!
[nltk_data]     | Downloading package pe08 to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package pe08 is already up-to-date!
[nltk_data]     | Downloading package perluniprops to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package perluniprops is already up-to-date!
[nltk_data]     | Downloading package pil to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package pil is already up-to-date!
[nltk_data]     | Downloading package pl196x to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package pl196x is already up-to-date!
[nltk_data]     | Downloading package porter_test to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package porter_test is already up-to-date!
[nltk_data]     | Downloading package ppattach to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package ppattach is already up-to-date!
[nltk_data]     | Downloading package problem_reports to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package problem_reports is already up-to-date!
[nltk_data]     | Downloading package product_reviews_1 to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package product_reviews_1 is already up-to-date!
[nltk_data]     | Downloading package product_reviews_2 to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package product_reviews_2 is already up-to-date!
[nltk_data]     | Downloading package propbank to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package propbank is already up-to-date!
[nltk_data]     | Downloading package pros_cons to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package pros_cons is already up-to-date!
[nltk_data]     | Downloading package ptb to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package ptb is already up-to-date!
[nltk_data]     | Downloading package punkt to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package punkt is already up-to-date!
```

```
[nltk_data]     | Downloading package qc to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package qc is already up-to-date!
[nltk_data]     | Downloading package reuters to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package reuters is already up-to-date!
[nltk_data]     | Downloading package rslp to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package rslp is already up-to-date!
[nltk_data]     | Downloading package rte to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package rte is already up-to-date!
[nltk_data]     | Downloading package sample_grammars to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package sample_grammars is already up-to-date!
[nltk_data]     | Downloading package semcor to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package semcor is already up-to-date!
[nltk_data]     | Downloading package senseval to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package senseval is already up-to-date!
[nltk_data]     | Downloading package sentence_polarity to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package sentence_polarity is already up-to-date!
[nltk_data]     | Downloading package sentiwordnet to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package sentiwordnet is already up-to-date!
[nltk_data]     | Downloading package shakespeare to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package shakespeare is already up-to-date!
[nltk_data]     | Downloading package sinica_treebank to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package sinica_treebank is already up-to-date!
[nltk_data]     | Downloading package smultron to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package smultron is already up-to-date!
[nltk_data]     | Downloading package snowball_data to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package snowball_data is already up-to-date!
[nltk_data]     | Downloading package spanish_grammars to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package spanish_grammars is already up-to-date!
[nltk_data]     | Downloading package state_union to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package state_union is already up-to-date!
[nltk_data]     | Downloading package stopwords to
[nltk_data]     |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]     |   Package stopwords is already up-to-date!
[nltk_data]     | Downloading package subjectivity to
```

```
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package subjectivity is already up-to-date!
[nltk_data]    | Downloading package swadesh to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package swadesh is already up-to-date!
[nltk_data]    | Downloading package switchboard to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package switchboard is already up-to-date!
[nltk_data]    | Downloading package tagsets to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package tagsets is already up-to-date!
[nltk_data]    | Downloading package timit to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package timit is already up-to-date!
[nltk_data]    | Downloading package toolbox to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package toolbox is already up-to-date!
[nltk_data]    | Downloading package treebank to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package treebank is already up-to-date!
[nltk_data]    | Downloading package twitter_samples to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package twitter_samples is already up-to-date!
[nltk_data]    | Downloading package udhr to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package udhr is already up-to-date!
[nltk_data]    | Downloading package udhr2 to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package udhr2 is already up-to-date!
[nltk_data]    | Downloading package unicode_samples to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package unicode_samples is already up-to-date!
[nltk_data]    | Downloading package universal_tagset to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package universal_tagset is already up-to-date!
[nltk_data]    | Downloading package universal_treebanks_v20 to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package universal_treebanks_v20 is already up-to-
[nltk_data]    |       date!
[nltk_data]    | Downloading package vader_lexicon to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package vader_lexicon is already up-to-date!
[nltk_data]    | Downloading package verbnet to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package verbnet is already up-to-date!
[nltk_data]    | Downloading package verbnet3 to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package verbnet3 is already up-to-date!
[nltk_data]    | Downloading package webtext to
```

```
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package webtext is already up-to-date!
[nltk_data]    | Downloading package wmt15_eval to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package wmt15_eval is already up-to-date!
[nltk_data]    | Downloading package word2vec_sample to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package word2vec_sample is already up-to-date!
[nltk_data]    | Downloading package wordnet to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package wordnet is already up-to-date!
[nltk_data]    | Downloading package wordnet2021 to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package wordnet2021 is already up-to-date!
[nltk_data]    | Downloading package wordnet2022 to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package wordnet2022 is already up-to-date!
[nltk_data]    | Downloading package wordnet31 to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package wordnet31 is already up-to-date!
[nltk_data]    | Downloading package wordnet_ic to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package wordnet_ic is already up-to-date!
[nltk_data]    | Downloading package words to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package words is already up-to-date!
[nltk_data]    | Downloading package ycoe to
[nltk_data]    |     C:\Users\murat\AppData\Roaming\nltk_data...
[nltk_data]    |   Package ycoe is already up-to-date!
[nltk_data]    |
[nltk_data]  Done downloading collection all

True
```

1. Зчитати файл text1.

```
with open('text1.txt', 'r') as f:
    text = f.read()

text

'Isa Whitney, brother of the late Elias Whitney, D.D., Principal of
the\nTheological College of St. Georgeâ€™s, was much addicted to
opium. The\nhabit grew upon him, as I understand, from some foolish
freak when he\nwas at college; for having read De Quinceyâ€™s
description of his dreams\nand sensations, he had drenched his tobacco
with laudanum in an attempt\nto produce the same effects. He found, as
so many more have done, that\nthe practice is easier to attain than to
get rid of, and for many years\nhe continued to be a slave to the
drug, an object of mingled horror\nand pity to his friends and
```

```
relatives. I can see him now, with yellow,\npasty face, drooping lids,
and pin-point pupils, all huddled in a\nchair, the wreck and ruin of a
noble man.\n\nOne nightâ€”it was in June, â€™89â€”there came a ring to
my bell, about the\nhour when a man gives his first yawn and glances
at the clock. I sat up\nin my chair, and my wife laid her needle-work
down in her lap and made\na little face of disappointment.'
```

а) Порахувати кількість речень в тексті;

```
sentences = nltk.sent_tokenize(text)
print(sentences)

sent_num = len(sentences)

print(f"\nКількість речень:", sent_num)
```

```
['Isa Whitney, brother of the late Elias Whitney, D.D., Principal of
the\nTheological College of St. Georgeâ€™s, was much addicted to
opium.', 'The\nhabit grew upon him, as I understand, from some foolish
freak when he\nwas at college; for having read De Quinceyâ€™s
description of his dreams\nand sensations, he had drenched his tobacco
with laudanum in an attempt\nto produce the same effects.', 'He found,
as so many more have done, that\nthe practice is easier to attain than
to get rid of, and for many years\nhe continued to be a slave to the
drug, an object of mingled horror\nand pity to his friends and
relatives.', 'I can see him now, with yellow,\npasty face, drooping
lids, and pin-point pupils, all huddled in a\nchair, the wreck and
ruin of a noble man.', 'One nightâ€”it was in June, â€™89â€”there came
a ring to my bell, about the\nhour when a man gives his first yawn and
glances at the clock.', 'I sat up\nin my chair, and my wife laid her
needle-work down in her lap and made\na little face of
disappointment.']

Кількість речень: 6
```

б) вивести 10 слів, які зустрічаються найчастіше;

```
from nltk.corpus import stopwords


stop_words = set(stopwords.words('english'))

words = nltk.word_tokenize(text)
filtered_words = [word.lower() for word in words if word.isalnum() and
word.lower() not in stop_words]

freq_dist = nltk.FreqDist(filtered_words)
for word, count in freq_dist.most_common(10):
    print(f"{word}: {count}")
```

```
whitney: 2
college: 2
many: 2
face: 2
chair: 2
man: 2
isa: 1
brother: 1
late: 1
elias: 1
```

в) провести лематизацію слів третього речення (попередньо визначивши частини мови).

```
third_sentence = sentences[2]
third_sentence
```

```
'He found, as so many more have done, that\nthe practice is easier to
attain than to get rid of, and for many years\nhe continued to be a
slave to the drug, an object of mingled horror\nand pity to his
friends and relatives.'
```

```
tokens = nltk.word_tokenize(third_sentence)
tokens = [token.lower() for token in tokens if token not in ',.!?']
tokens
# print("Частини мови:", pos_tagged)
```

```
['he',
 'found',
 'as',
 'so',
 'many',
 'more',
 'have',
 'done',
 'that',
 'the',
 'practice',
 'is',
 'easier',
 'to',
 'attain',
 'than',
 'to',
 'get',
 'rid',
 'of',
 'and',
 'for',
 'many',
 'years',
```

```
 'he',
 'continued',
 'to',
 'be',
 'a',
 'slave',
 'to',
 'the',
 'drug',
 'an',
 'object',
 'of',
 'mingled',
 'horror',
 'and',
 'pity',
 'to',
 'his',
 'friends',
 'and',
 'relatives']

pos_tagged = nltk.pos_tag(tokens)
print("Частини мови:", pos_tagged)
```

```
Частини мови: [('he', 'PRP'), ('found', 'VBD'), ('as', 'IN'), ('so',
'RB'), ('many', 'JJ'), ('more', 'JJR'), ('have', 'VBP'), ('done',
'VBN'), ('that', 'IN'), ('the', 'DT'), ('practice', 'NN'), ('is',
'VBZ'), ('easier', 'JJR'), ('to', 'TO'), ('attain', 'VB'), ('than',
'IN'), ('to', 'TO'), ('get', 'VB'), ('rid', 'JJ'), ('of', 'IN'),
('and', 'CC'), ('for', 'IN'), ('many', 'JJ'), ('years', 'NNS'), ('he',
'PRP'), ('continued', 'VBD'), ('to', 'TO'), ('be', 'VB'), ('a', 'DT'),
('slave', 'NN'), ('to', 'TO'), ('the', 'DT'), ('drug', 'NN'), ('an',
'DT'), ('object', 'NN'), ('of', 'IN'), ('mingled', 'JJ'), ('horror',
'NN'), ('and', 'CC'), ('pity', 'NN'), ('to', 'TO'), ('his', 'PRP$'),
('friends', 'NNS'), ('and', 'CC'), ('relatives', 'NNS')]
```

```
# Лематизація слів та вивід результату
from nltk.stem import WordNetLemmatizer
lemmatizer = WordNetLemmatizer()

lemmatized_words = []
for word, pos in pos_tagged:
    # Визначення частини мови для лематизації
    pos = pos[0].lower()
    pos = pos if pos in ['a', 'r', 'n', 'v'] else None
    if not pos:
        lemma = word
    else:
        lemma = lemmatizer.lemmatize(word, pos=pos)
```

```
    lemmatized_words.append(lemma)

print("Лематизоване речення:", ' '.join(lemmatized_words))
```

```
Лематизоване речення: he find as so many more have do that the
practice be easier to attain than to get rid of and for many year he
continue to be a slave to the drug an object of mingled horror and
pity to his friend and relative
```

Використати корпус Brown, перший текст категорії fiction.

а) Вивести перші 5 речень;

```
from nltk.corpus import brown

fiction_sentences = brown.sents(categories='fiction')[:5]
for sent in fiction_sentences:
    print(' '.join(sent), sep='\n')
```

```
Thirty-three
Scotty did not go back to school .
His parents talked seriously and lengthily to their own doctor and to
a specialist at the University Hospital -- Mr. McKinley was entitled
to a discount for members of his family -- and it was decided it would
be best for him to take the remainder of the term off , spend a lot of
time in bed and , for the rest , do pretty much as he chose --
provided , of course , he chose to do nothing too exciting or too
debilitating .
His teacher and his school principal were conferred with and everyone
agreed that , if he kept up with a certain amount of work at home ,
there was little danger of his losing a term .
Scotty accepted the decision with indifference and did not enter the
arguments .
```

б) Вивести 10 іменників, що зустрічаються найчастіше.

```
from collections import Counter

nouns = [word.lower() for word, pos in
nltk.pos_tag(brown.words(categories='mystery')) if
pos.startswith('NN') and word.lower() not in ["i'm", "mr."]]
nouns_freq = Counter(nouns)
most_common_nouns = nouns_freq.most_common(10)
print("10 найчастіших іменників:")
for adj, freq in most_common_nouns:
    print(adj)
```

```
10 найчастіших іменників:
man
```

```
time
door
car
room
way
something
office
eyes
hand
```