from nltk.tokenize import word\_tokenize

from nltk.corpus import stopwords

import glob

import nltk

import os

import pathlib

nltk.download('punkt')

# 1. How to read multiple text files from folder in Python?

def quest1():

    for root, dirs, files in os.walk('C:/Users/Ezra Muir/Documents/Training-Work/Python/Oct\_Learn/Oct28/'):

        for file in files:

            filename, extension = os.path.splitext(file)

            if extension == '.py':

                # print(filename)

                print(f'{filename}{extension}')

# 2. How to iterate over files in directory using Python?

def quest2():

    path = 'C:/Users/Ezra Muir/Documents/Training-Work/Python/Oct\_Learn/Oct28/'

    # os.chdir(path)

    print("\nUSING os.listdir()")

    for fileName1 in os.listdir(path):

        f = os.path.join(path, fileName1)

        # if os.path.isfile(f) and os.path.basename == '.py':

        if os.path.isfile(f):

            print(f)

    print("\nUSING os.scandir()")

    for fileName2 in os.scandir(path):

        if fileName2.is\_file():

            print(fileName2.path)

    print("\nUSING pathlib module")

    files = pathlib.Path(path).glob('\*')

    for f in files:

        print(f)

    print("\nUSING os.walk()")

    for root, dirs, files in os.walk(path):

        for fileName3 in files:

            print(os.path.join(root, fileName3))

    print("\nUSING glob module")

    for fileName4 in glob.iglob(f'{path}/\*'):

        print(fileName4)

# 3. How to get file extension in Python?

def quest3():

    path3 = 'C:/Users/Ezra Muir/Documents/Training-Work/Python/Oct\_Learn/Oct28/Assignment53.py'

    filename, file\_ext1 = os.path.splitext(path3)

    def getFile():

        head, tail = os.path.split(path3)

        return tail

    print("\nUSING splitect()")

    print("File:      ", getFile())

    print("extension: ", file\_ext1)

    print("\nUSING pathlib module")

    file\_ext2 = pathlib.Path('Assignment53.py').suffix

    print("Extension: ", file\_ext2)

# 4. Create Inverted Index for File using Python

def quest4():

    file = open('Assign53.txt', encoding='utf8')

    read = file.read()

    file.seek(0)

    read

    line = 1

    for word in read:

        if word == '\n':

            line += 1

    print(f"\nLines: {line}")

    array = []

    for i in range(line):

        array.append(file.readline())

    print(f"\n{array}")

    # REMOVING PUNCTUATIONS

    punc = '''!()-[]{};:'"\, <>./?@#$%^&\*\_~'''

    for el in read:

        if el in punc:

            read = read.replace(el, " ")

    read

    # to maintain uniformity

    read = read.lower()

    print(f"\n{read}")

    # CLEAN DATA BY REMOVING STOPWORDS

    for i in range(1):

        textTokens = word\_tokenize(read)

    tokensWithoutSW = [word for word in textTokens if not word in stopwords.words()]

    print(f"\n{tokensWithoutSW}")

    # CREATE AN INVERTED INDEX

    dict = {}

    for i in range(line):

        check = array[i].lower()

        for item in tokensWithoutSW:

            if item in check:

                if item not in dict:

                    dict[item] = []

                if item in dict:

                    dict[item].append(i+1)

    print(f"\n{dict}")

# 5. Write a program in Python to Append content of one text file to another

def quest5():

    # entering the file names

    firstFile = 'fir.txt'

    secondFile = 'sec.txt'

    # opening both files in read only mode to read initial contents

    with open(firstFile, 'r', encoding="utf8") as first:

        with open(secondFile, 'r', encoding="utf8") as second:

            print("\nBEFORE APPENDING")

            print('First File: ', first.read())

            print('Second File: ', second.read())

    # opening first file in append mode and second file in read mode

    with open(firstFile, 'a+', encoding="utf8") as first:

        with open(secondFile, 'r', encoding="utf8") as second:

            first.write("\n")

            first.write(second.read())

            first.seek(0)

            second.seek(0)

            print("\nAFTER APPENDING")

            print('First File: ', first.read())

            print('Second File: ', second.read())

# Calling Functions

def main():

    # quest1()

    # quest2()

    # quest3()

    quest4()

    # quest5()

if \_\_name\_\_ == "\_\_main\_\_":

    main()

OUTPUTS









