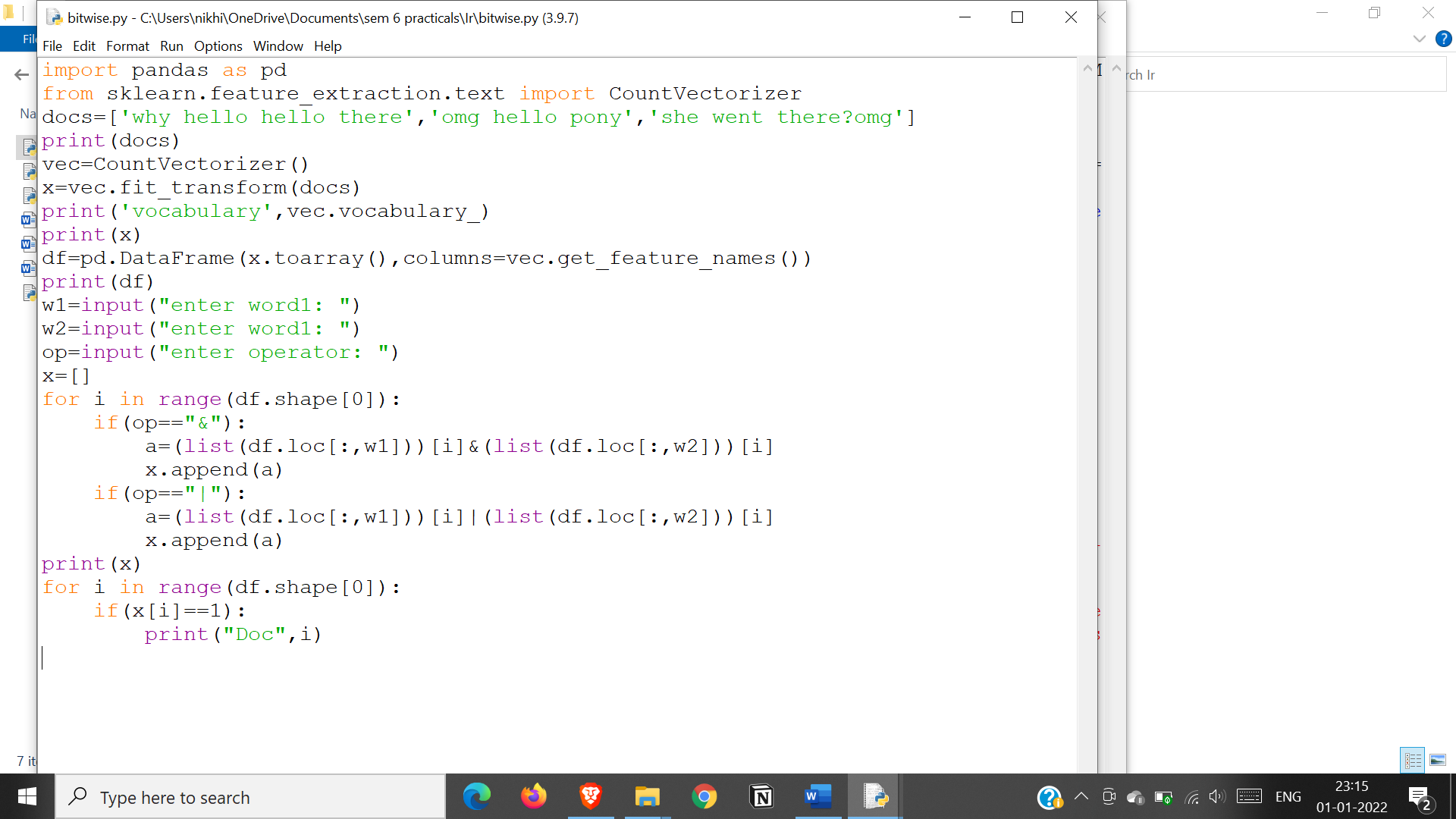
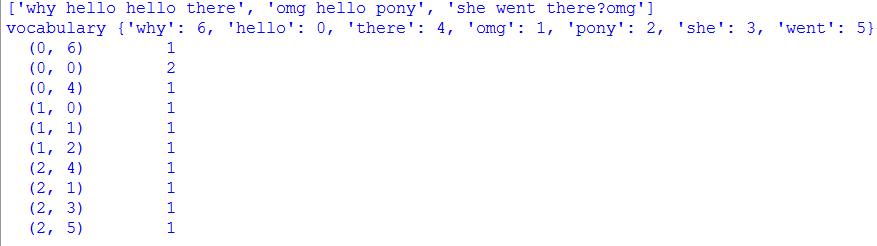
**Practical:1**

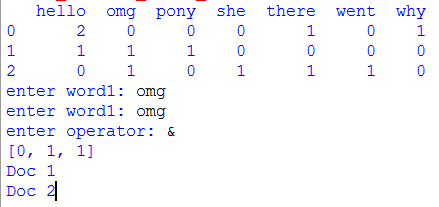
**Aim: Write a program to demonstrate bitwise operator**

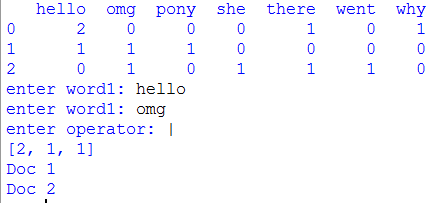
**Intro: CountVectorizer**is a great tool provided by the scikit-learn library in Python. It is used to transform a given text into a vector on the basis of the frequency (count) of each word that occurs in the entire text. This is helpful when we have multiple such texts, and we wish to convert each word in each text into vectors (for using in further text analysis).

****

**output:**

****

****

****

**Practical:3**

**Aim. Implement Dynamic programming algorithm for computing the edit distance between strings s1 and s2. (Hint. Levenshtein Distance)**

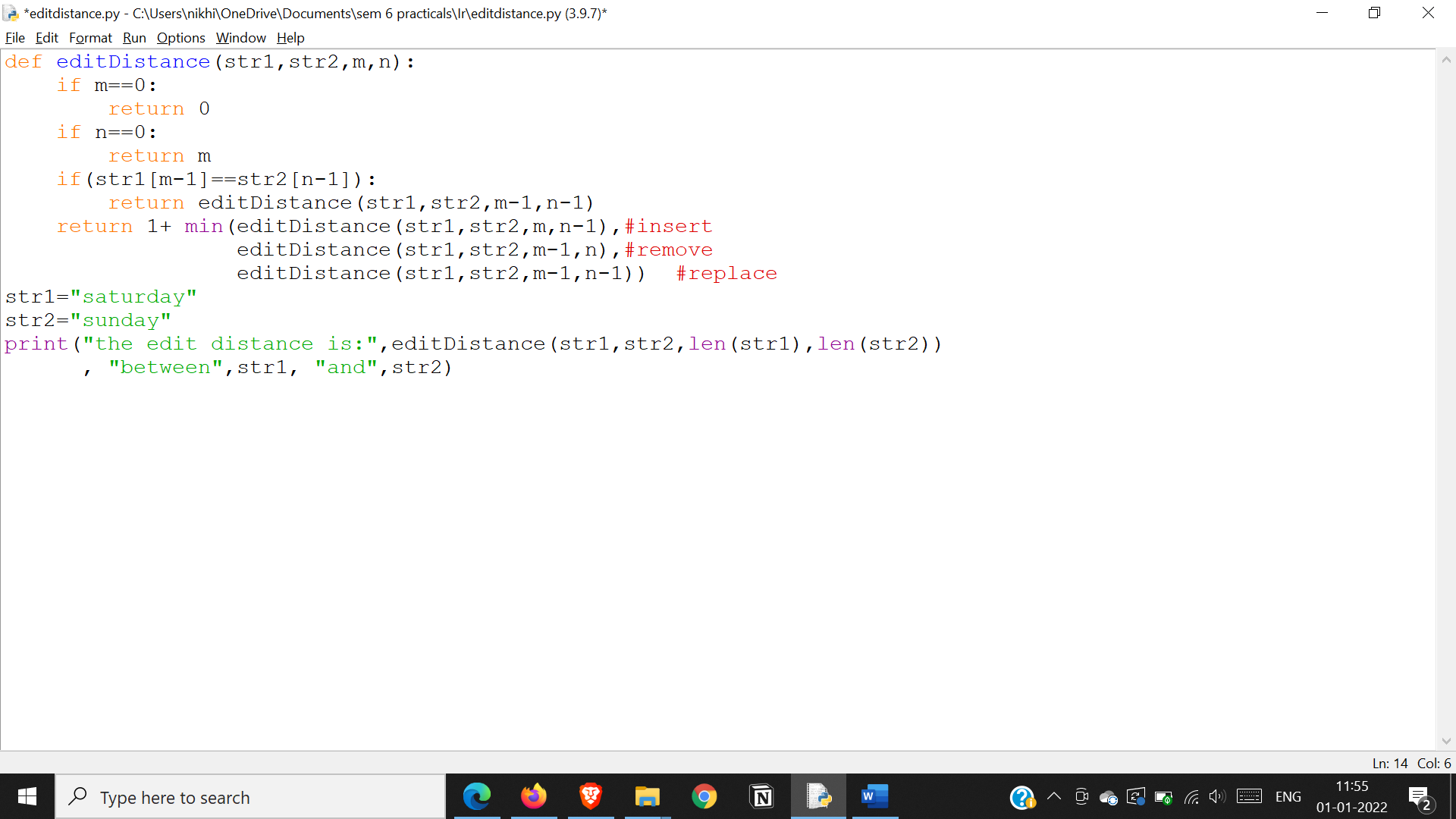
**Intro:**

Given two strings word1 and word2, return the minimum number of operations required to convert word1 to word2.

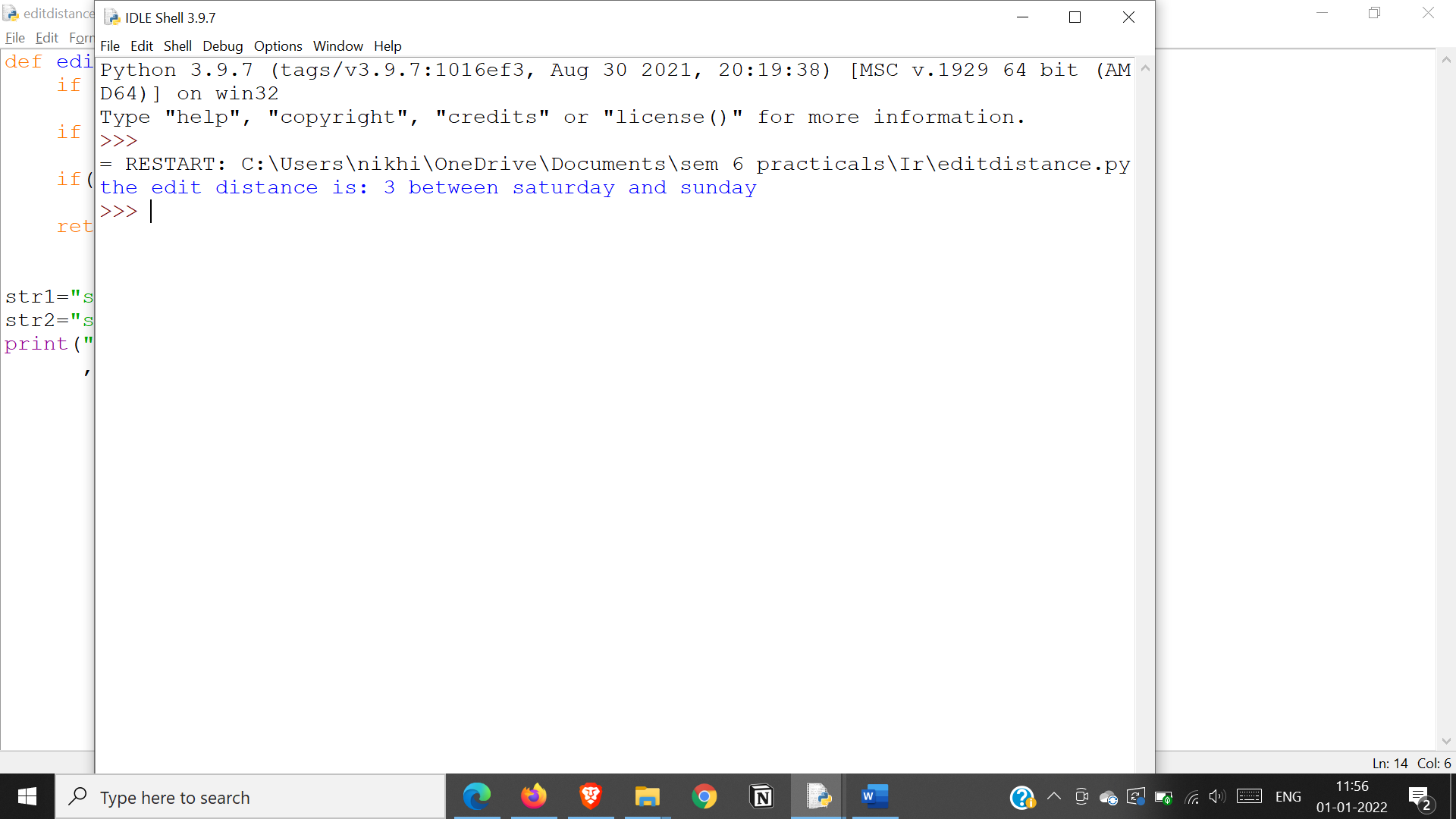
You have the following three operations permitted on a word:

* Insert a character
* Delete a character
* Replace a character

**Input:**

****

**Output**

****

**Practical:3**

**Aim: Write a map-reduce program to count the number of occurrences of each alphabetic character in the given dataset. The count for each letter should be case-insensitive (i.e., include both upper-case and lower-case versions of the letter; Ignore non-alphabetic characters).**

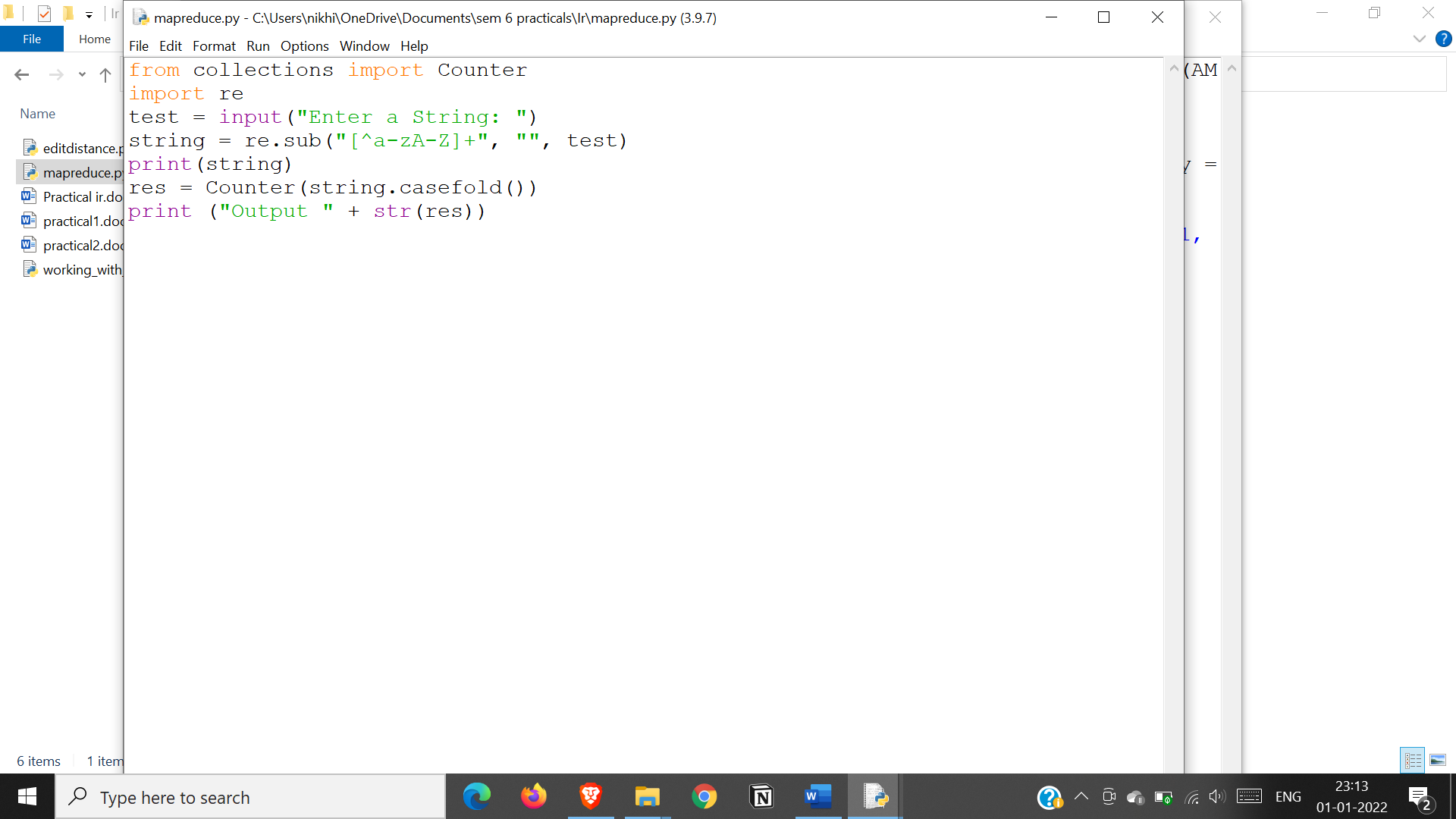
**Intro:**

A counter is a container that stores elements as dictionary keys, and their counts are stored as dictionary values.

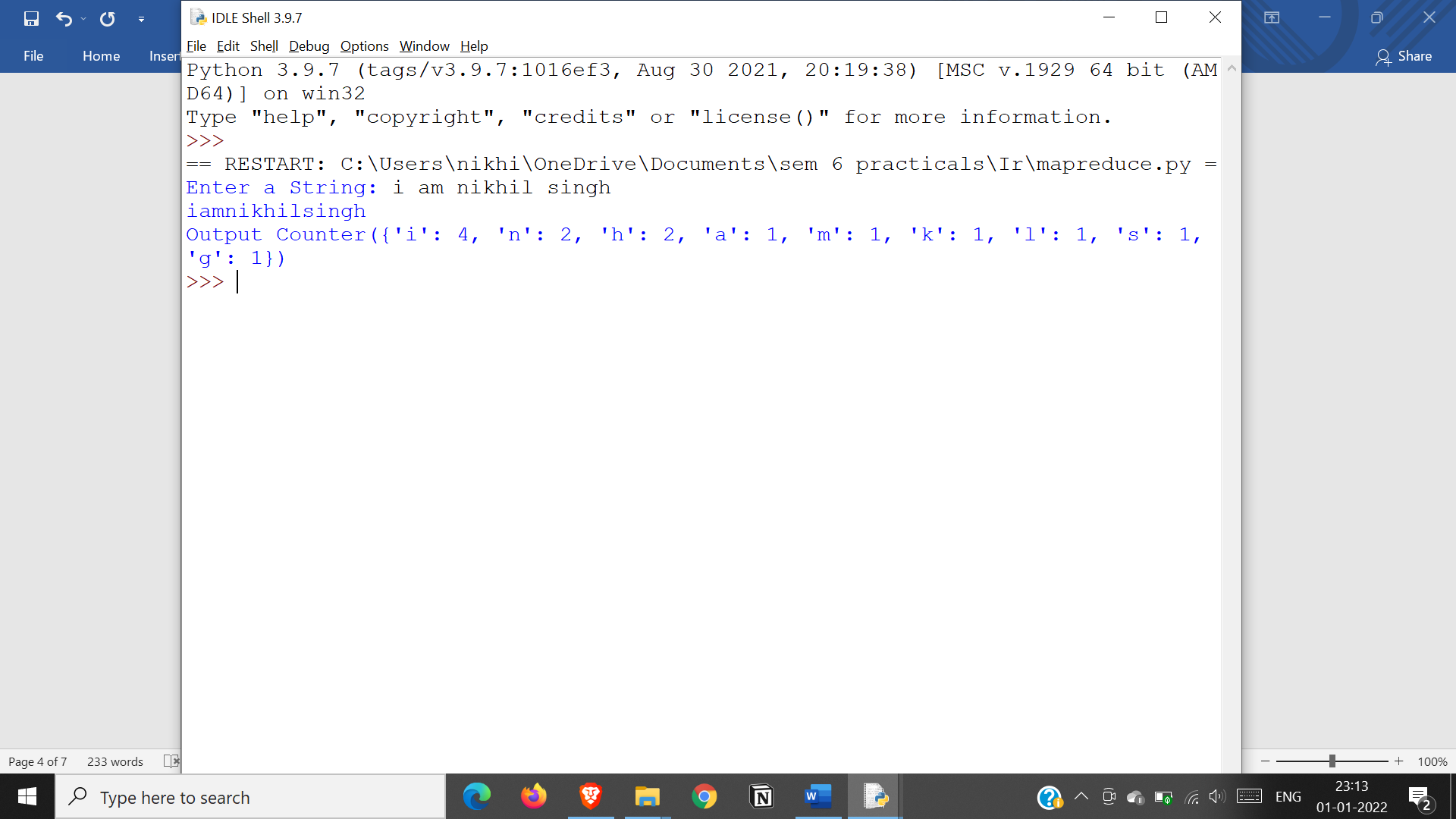
The casefold() method removes all case distinctions present in a string. It is used for caseless matching, i.e. ignores cases when comparing.

A regular expression (or RE) specifies a set of strings that matches it; the functions in this module let you check if a particular string matches a given regular expression (or if a given regular expression matches a particular string, which comes down to the same thing).

**Input:**

****

**Output:**

****

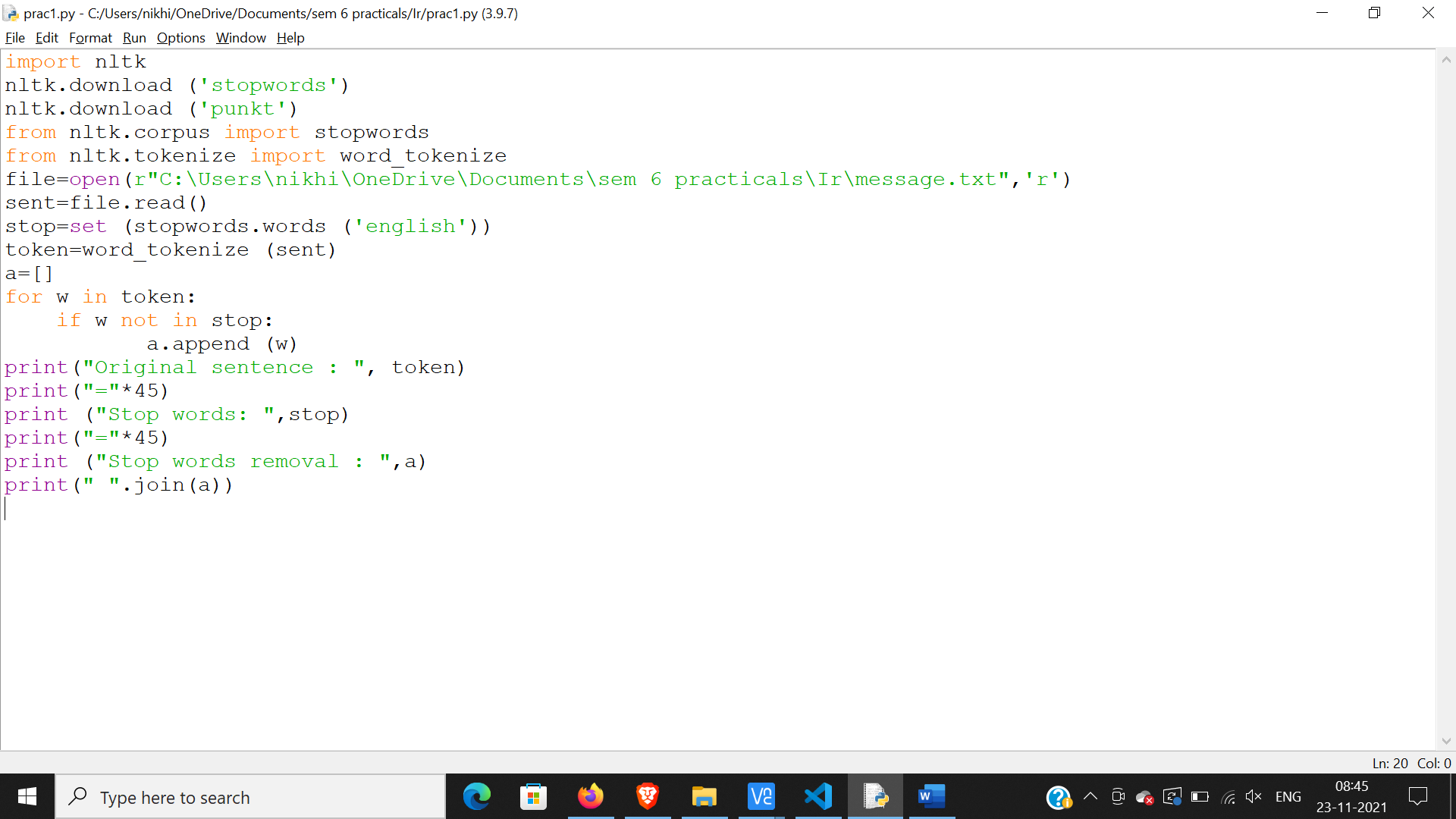
**Practical:7**

**Aim: Write a program for Pre-processing of a Text Document: stop word removal.**

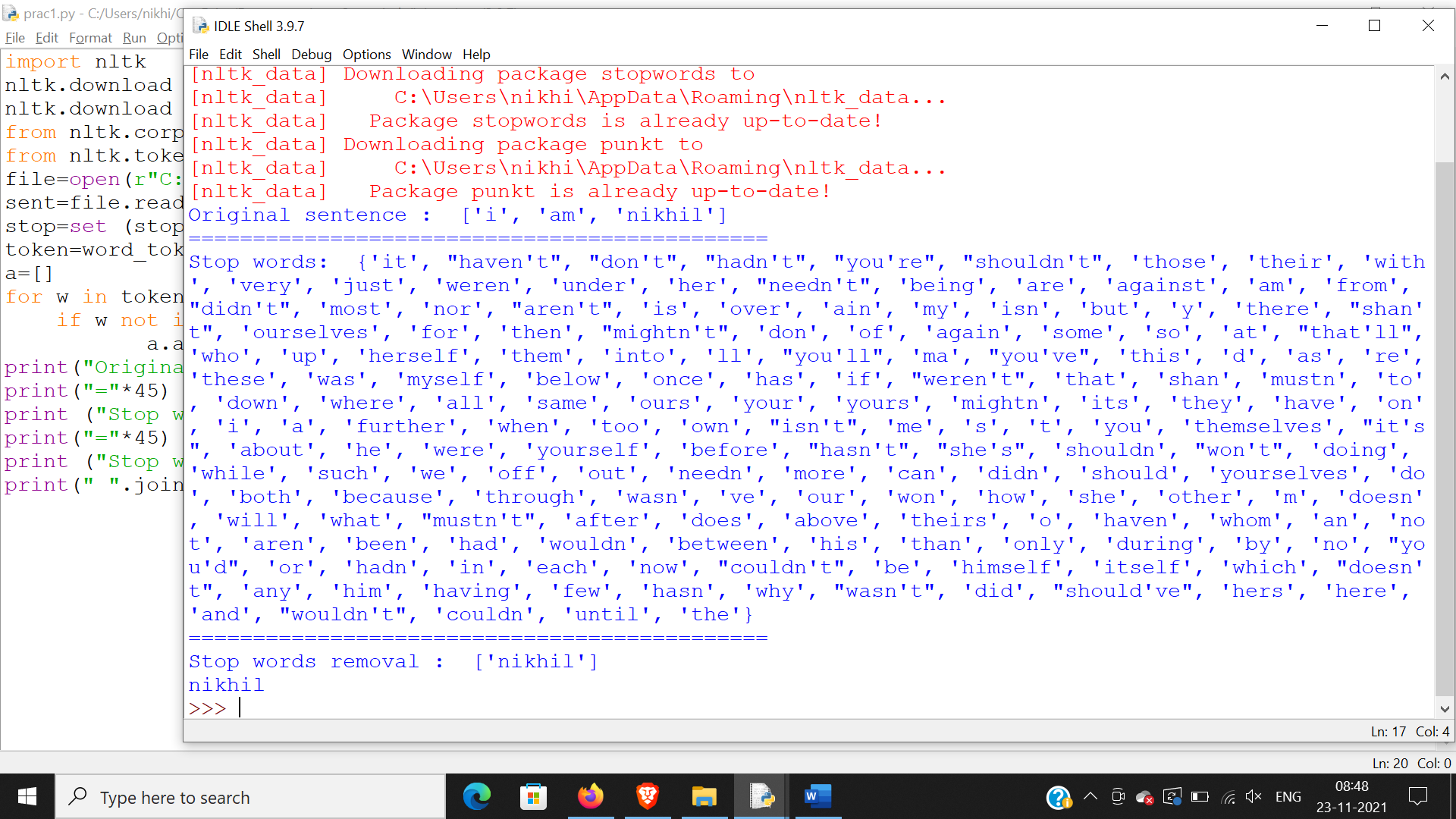
**Intro:**

A stop word is a commonly used word (such as “the”, “a”, “an”, “in”) that a search engine has been programmed to ignore, both when indexing entries for searching and when retrieving them as the result of a search query.   
We would not want these words to take up space in our database, or taking up valuable processing time. For this, we can remove them easily, by storing a list of words that you consider to stop words.

NLTK(Natural Language Toolkit) in python has a list of stopwords stored in 16 different languages. You can find them in the nltk\_data directory

****

**output:**

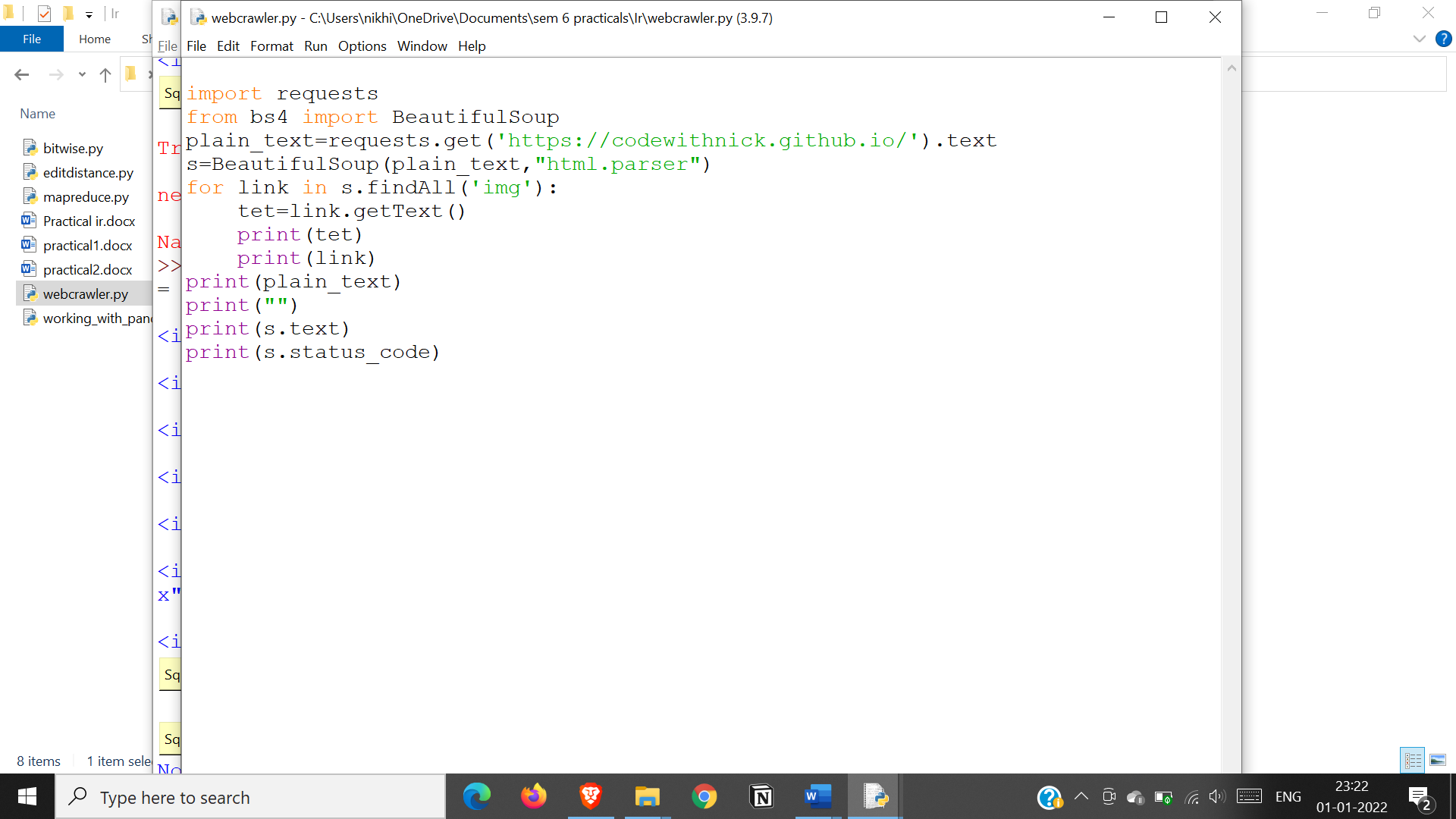
****

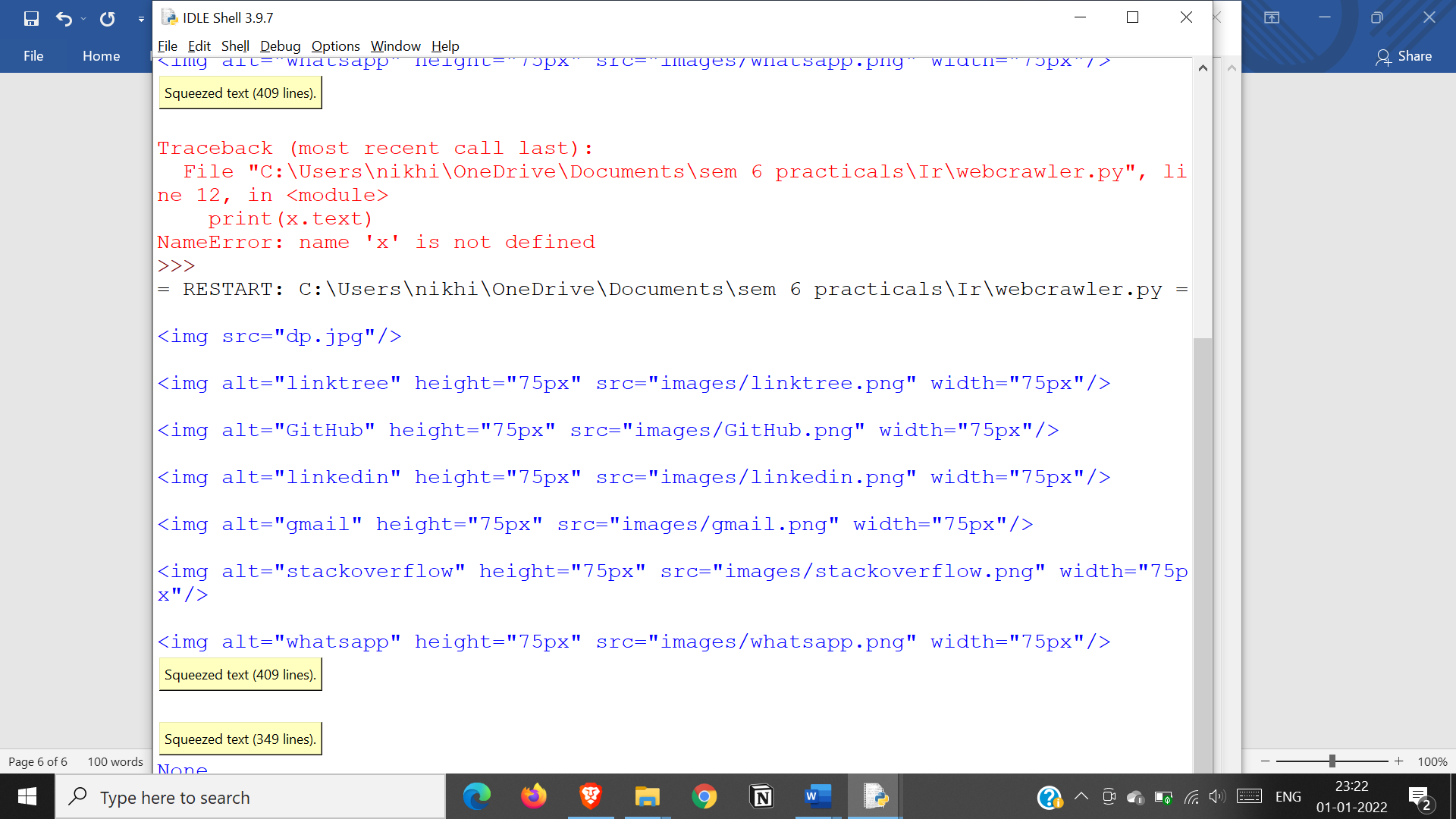
**Practical:9**

**Aim: Write a program to implement simple web crawler.**

**Intro:**

As the name suggests, the web crawler is a computer program or automated script that crawls through the World Wide Web in a predefined and methodical manner to collect data. The web crawler tool pulls together details about each page: titles, images, keywords, other linked pages, etc. It automatically maps the web to search documents, websites, RSS feeds, and email addresses. It then stores and indexes this data.

**Output:**

****