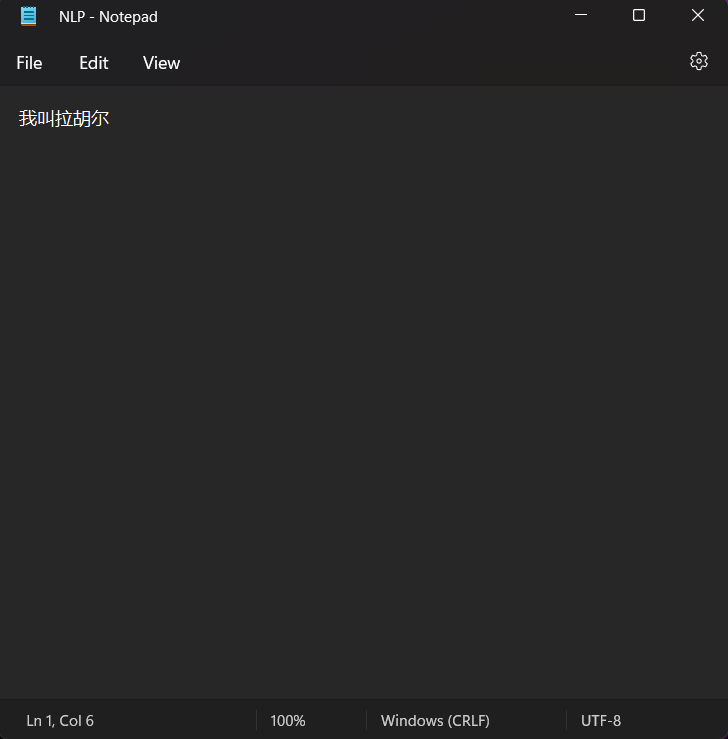
NATURAL LANGUAGE PROCESSESING

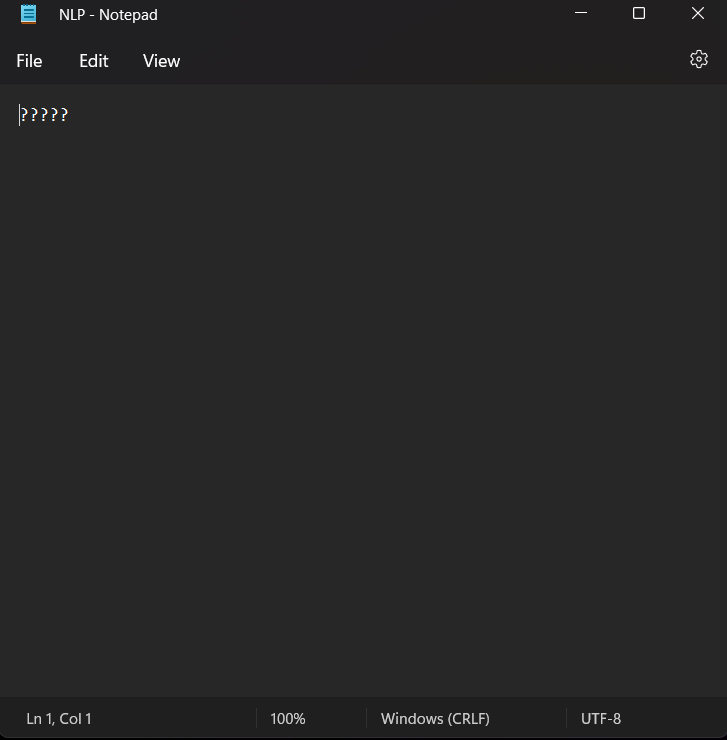
NAME: RAHUL GUDIVADA REG NO: 19BCE2469

DATE: 15/07/2022

**Task 1:**

Encoding:

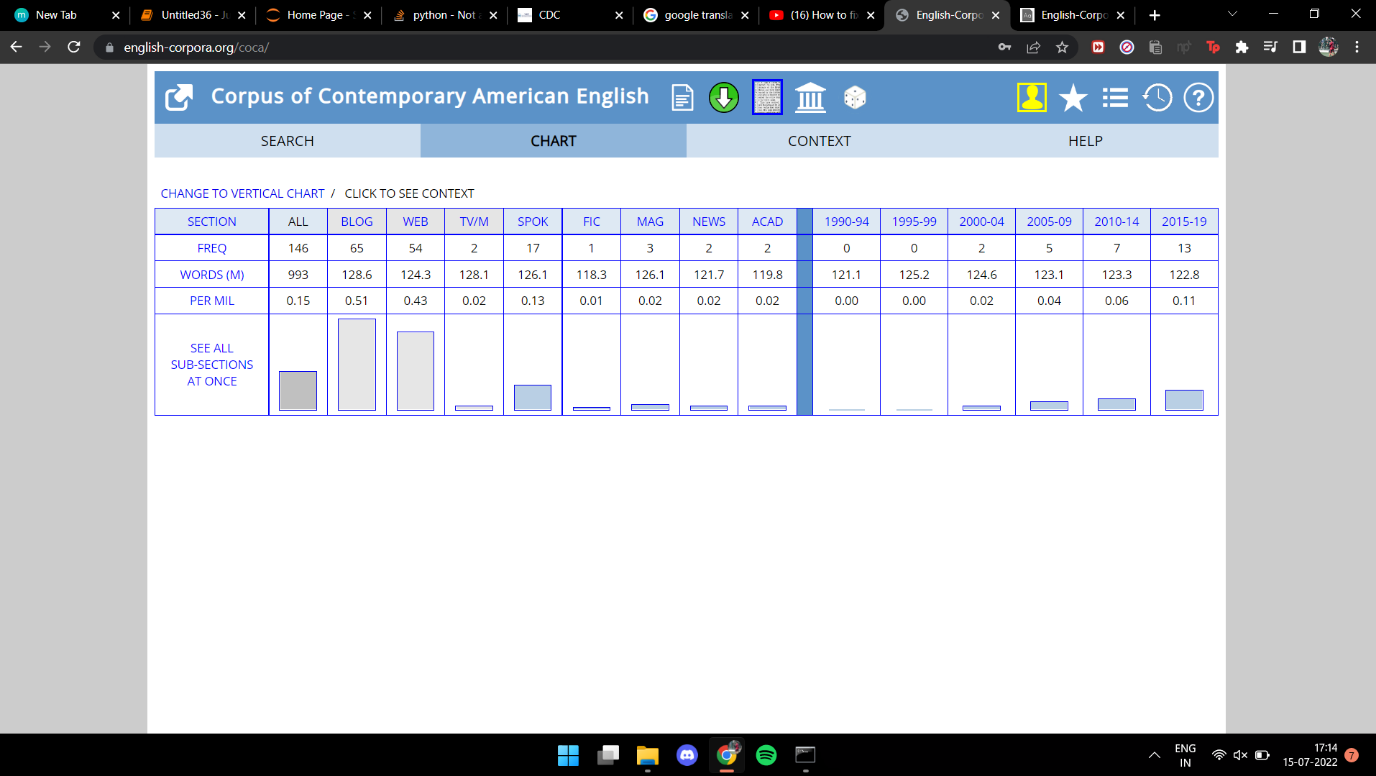




**Task 2:**

**2.1**

COCA, Story from Data



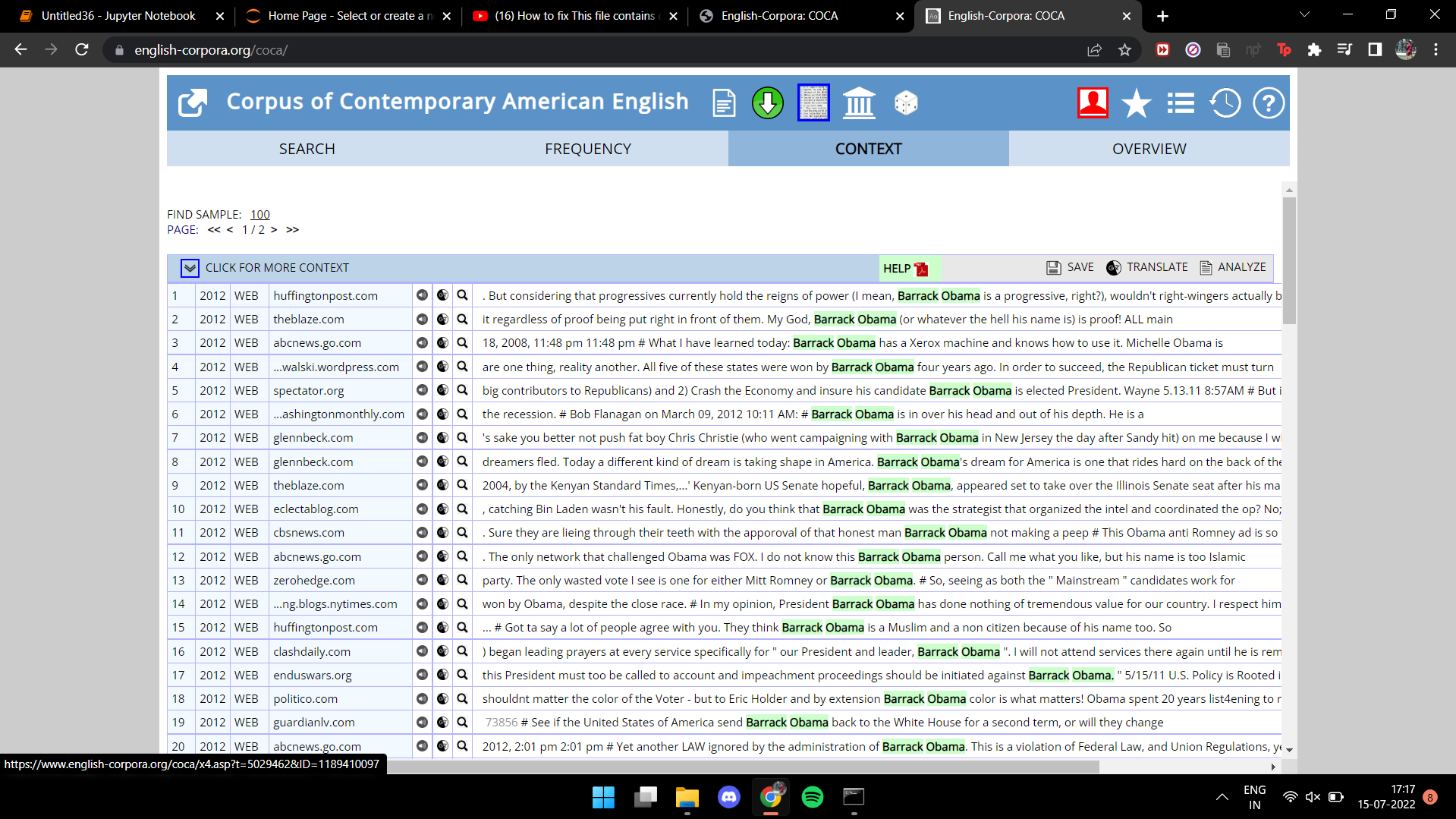
From the graph,we can clearly see that Barrack Obama wasn't famous during the time period of 1990-94 till 2000-2004. But slowly slowly he started gaining popularity as he started entering politics. From these graphs we can see that from 2005 onwards, people starting looking for him more in internet and how famous he became.

We can also see the frequency that is the frequency of him being in blogs is 65, whereas in WEB it is 54.

We can also see how many times his name as a WORD(M) is used in Blogs which is 65,wheres in WEB it is 54.

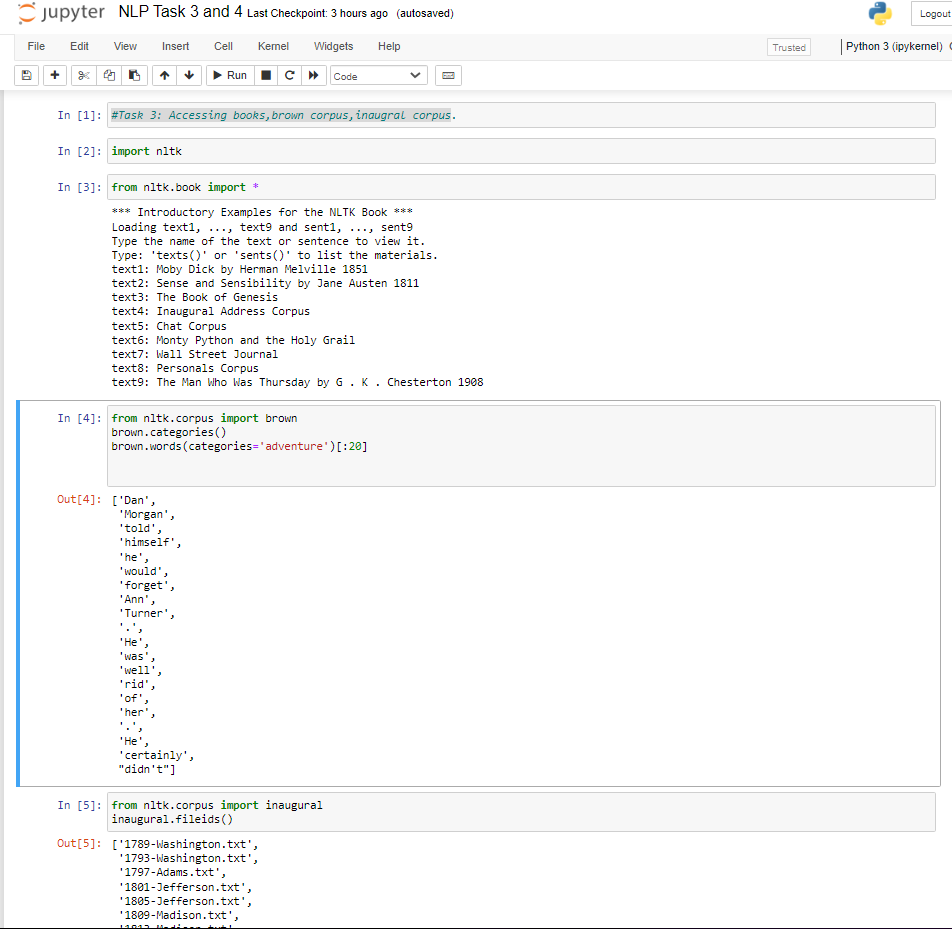
**2.2**

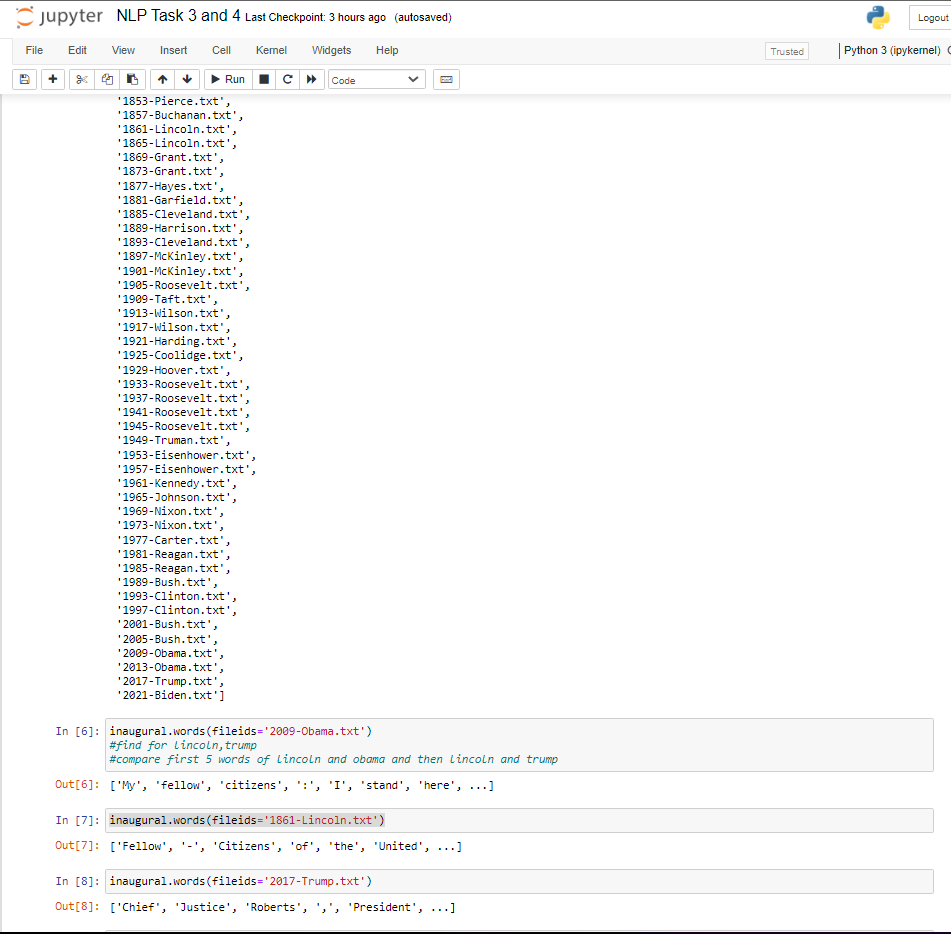
Concordance



**Task 3:**

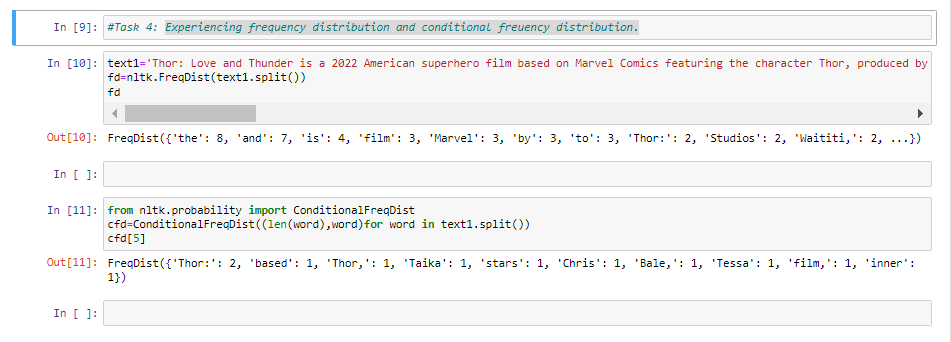
Accessing books, brown corpus, inaugural corpus





**Task 4:**

Experiencing frequency distribution and conditional frequency distribution.



**Task 5:**

Suggest a Corpus Application

**Corpus Approaches to Social-Media**

The language of online communities, linguistic variety in brief media texts, and the use of images in computer-mediated communication are all examined from the perspective of corpora. The collection's in-depth descriptions of the methodological aspects of working with social media corpora are one of its strongest features. The collection includes research using novel and creative research approaches for the analysis of multimodal material and atypical corpus texts, as well as research using conventional corpus linguistic methods to social media data.

Even if social media has become increasingly popular as a news source, there are still concerns about how easily it may disseminate rumors and false information. It has been challenging to systematically examine this phenomenon, however, because it is necessary to gather extensive, objective data and make in-situ assessments of its accuracy. A corpus called CREDBANK was created to fill this gap by methodically fusing computer and human computing. CREDBANK is specifically a corpus of tweets, topics, events, and related human credibility assessments.