

## NATURAL LANGUAGE PROCESSING: AUTHOR STYLOMETRY 08

So i will start with the code, i will start with by preparing the data so first i will prepare a dictionary name paper let me first import the nltk so nltk working or not let me check first so i will write in the ipython console import nltk so its you can see that its not throwing error it means it has been imported successfully before you start your coding you need to download all the constrain all the collections of nltk so for that i will write nltk dot download this function will open new ppi through which you can download all the collections of nltk this required for this stylometry analysis so i will just download everything so up to we download we will start with our dictionary thing so i will write take papers as the dictionary so the first author is i will name is Madison, Madison and the key is Madison and the values will be list of papers which he wrote so i will that is ten, fourteen, thirty seven, thirty eight, thirty nine, forty, forty one, forty two, forty three, forty four, forty five, forty six, forty seven and forty eight so these are the file names if you see the files the name is the first name is same for all the files federal list but after underscore there is the numbering so i am defining these numbers as a list here ok so for Madison he wrote these as i told in a last video that Madison wrote around fourteen papers so these are the fourteen papers which he wrote and then we will create the dictionary for Hamilton so i will put a comma and then i will write Hamilton and Hamilton wrote around fifty two papers so i will write those fifty two numbers which are present there one, six, seven, eight, nine so i have return the list for Hamilton after that after Hamilton comes the Jay so i will create the key and value payer so jay and the values will be jay wrote four papers so two, three, four and five and after jay we have shared papers of Madison and Hamilton so i will just write shared and the values and the list of values are value list are eighteen, nineteen and twenty so three papers they wrote so collaboratively and there are disputed papers which we, we are not sure of who exactly wrote this papers so i will just write this disputed sorry disputed and the value will be a list of some papers which are forty nine, fifty, fifty one, fifty two, fifty three, fifty four, fifty five, fifty six, fifty seven, fifty eight, sixty two and sixty three so these are the papers which are disputed and we really don't know who exactly wrote so we are doing to analyse for these papers exactly will see that which who who can write this paper by analysing the word length graph of all the authors which you have stated before the disputed papers. Ok so our data is the dictionary is created now i will create i will create a definition of reading files so what the idea is, let me write the definition name read files ok this will taken take one argument name as file name and what we are going to do, we create another dictionary and this dictionary what it will contain. It will contain the key value as the authors name and the key will be the authors name the value will be whatever he wrote will keep all the data in a particular list as a string so the key will be the authors name and the value will be all the data, whatever he wrote so all the data as a single string so let me create a list of name string. So strings as list ok i will create my loop as for file and file name, the file name is my argument. With open i will open my file as so my file name is federal list and with an underscore now will have the curly braces so that i can put the file name so what, what exactly is this? It is the key value which the corresponding value for each key which for each author. So i will put it here and dot txt so i will created the

string for the file name so that i can open the exactly file exact file open of particular author. I will open it as f and i will append string which the strings are the list which i created the name of the list is strings dot append and what i am going to append? The file which is which i opened as f, i will read this f dot read and whatever it is reading i will just append it into this string list ok and i will just i will just join it and return it so i will just and what i am going to return is slash n dot join the string that's it. So what i am going to do with this function, each time i will for each author i will send the file name file name is the list of the which contain the numbers and through that these numbers i will read the file which is present in the data folder for the corresponding author and i will get all the data for particular author, read the string and put it in a single list. Ok what i am going to do with the single list will see in the next video. Thank you.