Project: Text Analysis/ Mining US President Speech

Balaram Tripathy

Contents

Problem 2: In this particular project, we are going to work on the inaugural corpora from the nltk in Python. We will be looking at the following speeches of the Presidents of the United States of America.

*Here we consider each speech as separate Text Data.

2.1. Find the number of characters, words and sentences for the mentioned documents.

Ans 2.1.

Text Data: Speech of President Franklin D. Roosevelt in 1941

No. of characters (with space) in Roosevelt Data = 7571

No. of characters (without space) in Roosevelt Data = 6174

No. of words (without space) in Roosevelt Data = 1360

No. of Sentence in Roosevelt Data = 68

Text Data: Speech of President John F. Kennedy in 1961

No. of characters (with space) in Kennedy Data = 7618

No. of characters (without space) in Kennedy Data = 6202

No. of words (without space) in Kennedy Data = 1390

No. of Sentence in Kennedy Data = 55

Text Data: Speech of President Richard Nixon in 1973

No. of characters (with space) in Nixon Data = 9991

No. of characters (without space) in Nixon Data = 8122

No. of words (without space) in Nixon Data = 2028

No. of Sentence in Nixon Data = 69

2.2 Remove all the stop words from all the three speeches.

Ans 2.2.

No. of Stop words in Roosevelt Speech Data = 632

No. of Stop words in Kennedy Speech Data = 618

No. of Stop words in Nixon Speech Data = 899

2.3 Which word occurs the most number of times in his inaugural address for each president? Mention the top three words. (after removing the stop words)

Ans 2.3.

Note- Before we reach to this step. We have removed the stop words, special characters and symbols. And we have done Stemming.

Most frequently used words after removing the stop words by Roosevelt are following,

Us	25
let	22
new	15
i	12
peac	11
we	9
great	9
america	9
nation	8
world.	8

Most frequently used words after removing the stop words by <u>Kennedy</u> are following,

Let	16
us	11
side	7
new	7
pledg	7
power	7
ask	6
nation	6
i	5
to	5

Most frequently used words after removing the stop words by <u>Nixon</u> are following,

let	16
us	11
side	7
new	7
pledg	7
power	7
ask	6
nation	6
i	5
to	5

2.4 Plot the word cloud of each of the speeches of the variable. (after removing the stop words)

Ans 2.4.

Word cloud of Roosevelt's Speech

Word cloud of Kennedy's Speech

Word cloud of Nixon's Speech