5/7/2 4 , 6 :3 1 PM 7.ipynb - Colab

<https://colab.research.google.com/drive/1> Ivo p TBNKKM0 Sm3 DGnV\_9 G9 8 p KNl h d vSS#p ri n tMo d e =true 1 /3

! sa**ipnlt**k

8)( f: bh**o3.1**0**/**-**pc**g**s** qmR**u i edyar nltk** 87: **f( )** bh**o**3**.1**0**/**-**p**g**s dya**q**met**R**ur n lick** 4: **f( )** h3**.10/**-**pck**g**s dya**q**met**R**ur n** j**olib** x8=> **2**5**( f**: **)** bh**o3.10/**-**pckgs** q**mt**R**uin edyalr** 462: **f( )** bh**o**3**.**10**/**-**pck**g**s dyal qmet**R**ur in**

tmopri nitk

tkw**nlo**a**d**( punkt**'** )

tkw**nlo**a**d**(wordnet**'** )

tkw**nlo**a**d**( download *(* vdc**p**on**tager'** )

tkw**nlo**a**d**( tpw**o**rd**s'** )

from motipr sent—tokenize

from motipr word—tokenize

from motpri stopwords

r**/**. **\_**][ ec**t**u**kp** Dw**load**i**ng** o**-**! \_][ gcP **ed**y**al**r **tunkp** is **/**. **\_**][ **re**c**k**p **t** D**wload**i**ng** ][ **k** Dw**li** v**dcpon\_tager** ro**/**. **\_dtklna**][ s/vco**er. \_dt**kl**a**][ U**zping /**. **\_**][ ec**k tprs** D**wload**i**ng** c/w**ors.** \_**dt**kl**a**][ U**zpin**g

from motipr

te1csn= = u" **i**s **05 wlk d**na **mo**r**e I** will walk 500 miles and I would walk 500 more. J k **u**d il**s** f **man**,**wh e**b **to** to be the man,who walks a thousand miles to t !" yur **a**l **do**wn

=st**nce**2 = fu dI" w **lyr as the p** played the play playfully as the players were fus"g**enplay**w**ith** in the playwith playfullness

Tokiz#t**nces**

sc1wr**dtoknize**= okenized\_sentencel = ( st**nce**1 )

wr**okizd**s**tn**c**e**2= = ( st**n**c**e**2 )

okzd#r**i**P**tnces Print tokenized sentence**

trinp ( okizds**tnce**1 )

trinp ( okizds**tnce**2 )

trinp ( wr: t1**s'** T**o**kniz**ed** Tokenized ,wrd\_t**o**kniz**e** (t**nce**1s ))

trinp ( \1 **t**cs:**'** Tok**n**iz**e**d Tokenized ,\_to**e**knizs (te1csn ))

trinp ( wrs: \2**'** T**o**k**n**iz**ed Tokenized** ,wrd\_t**o**kniz**e** (t**nce**2s ))

trinp ( \2 tcs:**'** Tok**n**iz**ed** Tokenized ,\_to**e**knizs (te2csn ))

**k** J**u 05** b r **tdan meils I**[ **,**. **who'
  
wfu**] **r** gd**teh ypl** I[ **in as,'
  
05 ml** J**u a**b **wr**: h [ **I ,**.**'** T**okni**z**ed ts**1

r. J[**'** c: f **05 u l** T**ki**z**d I e**b **mn,wh a o ts**1
  
fu : **r** gI[ **s,** T**o**kz**ed n**2 **playwith'**I[ c: **fu'**] Tokz g **with dn**2 **players**



#POSTain**g** Taggin

from motipr

wrd**tokn**iz**e**= **oken =** ( te1csn ) ( te2csn )

ed=**t**o**ag**sp = (token )

trinp ( Tgin Spech:**"** fo P**a**rts Parts of ,taged )

k **w** VB M dN05 m**lCD R) S**p**e**ch: **I,(**[**'** fo **P**rt**s** T**aing**



#Stp-Wrds1Removal **Stop-Words Removal 1**

from motpri stopwords

=tpw**o**rd**s** = tes ( tp.**words** ( english**"** ) )

trinp ( tpw**o**rd**s** )

wrd**tokn**iz**e**= **=** ( te1csn )

=cladtok**en** = [ ]

for in token:

fi f word otn ni

cleaned \_token, append (word)

trinp ( Tok St**nce**:**'** ,token )

trinp ( Cdl\a v**e**rsio**n**:**'** ,clad\_tok**en** )

f r mo **y eb** {**c u**p l **ha isn't",
  
w**r **mil** J**us da** b h **05** Stc: **I**[ **,**. T**oken '**

5/7/2 4 , 6 :3 1 PM 7 .i p yn b - Co l a b

**wk** J **05 tu**h **mI**[ vr**sio**: **,**.**'** C**laned**

#Stp-Wrds Removal 2

from motpri

twod=rsp tes tp.**words(** english**"** )

trinp \_wtodp**s**r)(

wrdiz(sc2) **token** = = cladtok**en** [] for i

fi otn i

trinp ( Tok St**nce**:**'** token) ,

trinp ( Cdl\a v**e**rsio**n**:**'** , clad\_tok**en**)

f r mo **y eb** {**c u**p l **ha isn't",**fu **wr** g**y**d**h** I[ Stc: **ias, pl'** Tok**en**

wth **fu**] g**yp**I**,**[**'** v**rsi**o: C**laned**



#Steming 1

from motipr

PoS() = s**tmer** wrdiz(sc1) **token** = wo**r**)([. **tmeds** = for i

trinp ( " otm**e**djins().

trinp ( u " **05 wlk mo**r**e d**na **i** j**us**d f k **an m**i**l wh to e**b "! ta yu **r do**wn l )

|  |  |
| --- | --- |
| , f j **hs** . **m lk wn 05 i a** y**u t dor e**b !  f j **k s ml** . **h wn 05 a e**b **i** y**u dr to** !  #Lemz**ati**on 1  from motipr  WodNL() = l**matizer** wrdiz(sc1) **token** = w**r**)([. **tadoemzl**p**ui** = trinp ( " m**e**d**ot**anzljp**ui**().  , J f **s i** . **m lk h wn 05 a** I y**u t dor e**b !  #Steming 2  from motipr  PoS() = s**tmer** wrdiz(sc2) **token** = wo**r**)([. **tmeds** = for i  trinp ( " otm**e**djins(). | for i |

xmplE # **a**g**e**su

ex**t**= " w**h t i**n s **p layer** wit fu" **yalp** e**h**

l.wrd**okn**iz(**e**x**t**) =

wo**r**)([. **tmeds** = for i

trinp ( " otm**e**djins().

n s **with** fu **player
  
wth** n s **i p layer**

|  |  |
| --- | --- |
| #Lemz**ati**on 2  from motipr  WodNL() = l**matizer** wrdiz(sc2) **token** = w**r**)([. **tadoemzl**p**ui** = trinp ( " m**e**d**ot**anzljp**ui**(). | for i |

puI **t**e**ncs** #

tespunci = fu I" w**th** d **player** w fus" g **in the ylap**

|  |  |
| --- | --- |
| h mazL pu**i tencs** #  w**r**)([. **tadoemzl**p**ui** =  h madzl **e**cs # **t**r**in**P  trinp ( " m**e**d**ot**anzljp**ui**().  g d I **with funs player  w** gd **th** I **i funs player** | for i |

h ttp [s://co](file:///s://co) l a b .re se a rch .g o og l e .co m/d ri ve/1 Ivo p TBNKKM0 Sm3 DGnV\_9 G9 8 p KNl h d vSS#p ri n tMo d e =true 2/3

t

5/7/2 4 , 6 :3 1 PM 7.ipynb - Colab

\* Instantiate the TfidfVectorizer fr i

d~~**~~**ata = : [1 ~ ~ **~**~ **05 mo**r**e ~~~ and ~** will walk 500 miles and I would walk 500 more. J to ␣ ~~ ~ ~**~~ ~~ ~~ man** man who walks a thousand miles to ~~ ~~ ~**~** ~~~ **~~**

<https://colab.research.google.com/drive/1> Ivo p TBNKKM0 Sm3 DGnV\_9 G9 8 p KNl h d vSS#p ri n tMo d e =true 3/3

t ext\_tf = tf.fit\_transform(data[ ]

p (

e e e e e e e e e e e e e e e e e e e e