М	T	W	T	F	5	S
Page No.					VO	UVA
Date					, .	

CHRIS SUNNY THA	LAVATA
	CIANIA
QUIZ NLP	
Dala:	
· docs, doc 2	doc 3, doc4, doss
these ave	
0	- A
wa, wa, wa	11 11 11 11 11 1 1 1 1 1 1 1 1 1 1 1 1
words	0
	650
- W6, W7 -	Stop words
Represent Corpus	10) pol = (0) 901
→ 1 Binary	
	ed = all Recott to the
Lip word progent	
n document, else	
	> Column: Repressents
(+)=-	La corde
To Total 2 -4 2 -4 2 -4 2 -4 2 -4 2 -4 2 -4 2 -	12 2 10 40, - 57
TF-IDF(+,d) = TF(+,d) XIDF(4) 17
IDF-TF (Inverse Doc	
	a low to be a low
Lo Histogram of	words in all decomends
0	different words
	(
ROW CLONELINGUE Ray	of words / Spip- Gram

14:

> two Create poin of words

(Target words & context words)

> Skip Gram: predicts the context

based on corrent word & surrounding words

Binary	Madrisc	1.71	politic.	· · ·		udo.	
				1 - 1	- / ,	1	
Document			W3	W4		W6	wy
doc 1	1.1	1.1	11	J.O. P.	,0 .	0	1
doc 2	1	190	1.11	. 6	10	0	4
doc3	1	0	0	0	1	6	1
doc 4	0, ,,,	0,	0	10/1/10/10	4	1	O
do 5	6	0	0	- Liver	0	1	0
	100	4.87	· ·	ile ,	а		
11)F(t) =	log (n/d (1	(((-	*Jr.	as in	
			dove pro				
	N-: Tobal						
,	dett)				ith do	ermi -	(,, "
	21 N 3 4 2 1 7 -				· Vac arr		67
	de (cos 1):				1	6	, .
	OF CWA						
	Acres and the second se				(110)		
	F-IDF (OV DOCT		TOP .	(MI)	h 27	
	1	. 5 /.			2		
		N 9 40.		23, 200	1	1 1 1	1 +
	lf (wa)						
June V	of (w2) "	F Colove	20	more	35 617		
d	bebu b	190 30 July	· · ·	the cast	Ji.		
Irras p -	7140	1:00	10 pp8	2001	11/10 1	·w3	80
	,						

by thistog

118 1 3 3 1 3 3 1 1 1

1361

6

Page No			
		YO	UVA
Date		3.55	

Cal				
col	1601	ah	6.0	
		-0	CON	18

	_	TF (doca)				
L NM	F(dal	TF(docz)	TF (de	TF(docg) DF	IDE
wa	1		١	0 0	3	1009(5/3)
wa.	\	1	0	0 0	2	109 (5/2)
Wa	-10/ 12	1 10 2	. 40	00-0	31 2032	
WS	0	0	6	4 0		log (S(1)
W6	0	0	1		2	109 (92)
w7	Λ	1	0	1	-	109 (5/2)
				, O. J.	3 -1	109(5/3)

TE IDE

-

-

-

0				
Document	Wi	·W2	wa wa	JUS W6 W7
dai	0.51	0.90	0.90	0 0 0.81
doca	6.81	0.90	0.90 6.	0110 0.81
" 3	0.51	0	0 0	090 0 0.51
· •	0	0	0 (.61	0.90 D.90 O
" 5	6	0	;0	0 0.90 0
			10	

CBOW - SKIP Gram

a) Tokenization: Break dodoment into words
Remove Stop words

6) Windowing: window size (2-3 words)

CBOW (Continous Bag of words)

Ly input: words within the window

L> output : Tanget words

Ly model: Surrounding Contexed

Page No YOUVA

Skip Gram: imput: Tanget would Context words Example with doc2 with window Size 2 Document 2: (we, wx, ws; w) Training Examples Predict W2 given WT 11