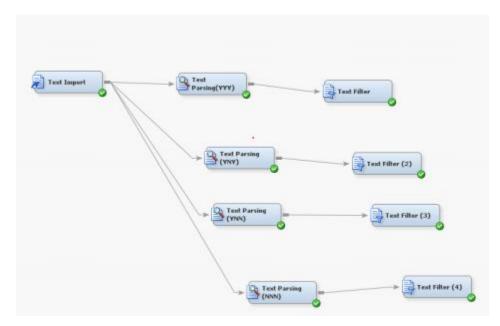
# SAS Part

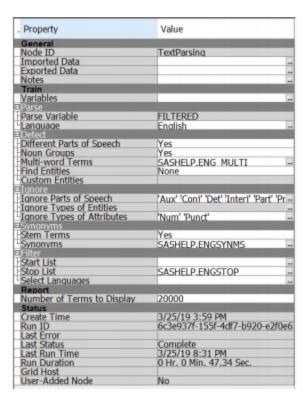
Q1



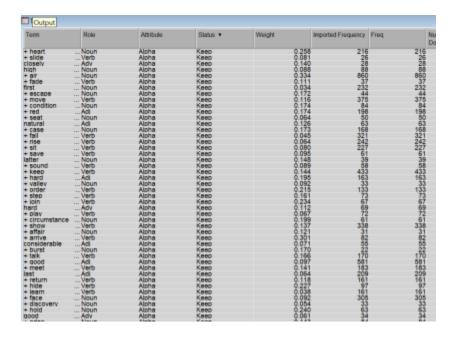
### 2. Import Property

. Property	Value
General	
Node ID	TextImport
Imported Data	
Exported Data	_
Notes	
Train	
Import File Directory	D:\Tamu\EM Proiects\HW9\Dat =
Destination Directory	D:\Tamu\EM Proiects\HW9\Wo =
Language	Enalish -
Extensions	22767
Text Size	32767
Web Crawl	
URL	2
-Depth	Restricted
Domain User Name	Restricted
Password	
Status	
Create Time	3/25/19 3:48 PM
Run ID	446311df-72ee-42f3-8f57-f69e7a
Last Error	11031101 7200 1213 0137 103070
Last Status	Complete
Last Run Time	3/25/19 8:31 PM
Run Duration	0 Hr. 0 Min. 2,25 Sec.
Grid Host	
User-Added Node	No

a. Stop Words-Yes; POS-Yes; Stem-Yes



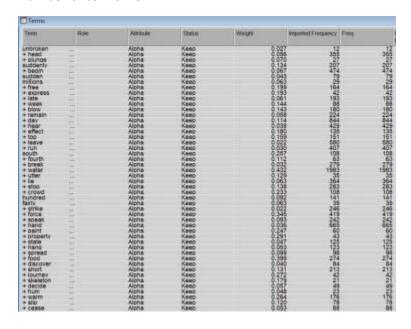
#### Number of terms = 2500



b) Stop Words-Yes; POS-No; Stem-Yes



#### Number of terms=4944



c) Stop Words-Yes; POS-No; Stem-No

Property	Value
General	
Node ID	TextParsing3
Imported Data	and the same of th
Exported Data	-
Notes	
Train	
Variables	-
Parse	
Parse Variable	FILTERED
Language	English
Detect	
Different Parts of Speech	No
Noun Groups	Yes
Multi-word Terms	SASHELP.ENG MULTI
Find Entities	None
Custom Entities	
Ignore	
Ignore Parts of Speech	'Aux' 'Coni' 'Det' 'Interi' 'Part' 'Pr
Ignore Types of Entitles	
Ignore Types of Attributes	'Num' 'Punct'
Synonyms	
Stem Terms	No
Synonyms	SASHELP.ENGSYNMS
Filter	
Start List	-
Stop List	SASHELP.ENGSTOP
Select Languages	-
Report	
Number of Terms to Display	20000
Status	
Create Time	3/25/19 8:39 PM
Run ID	55750c3e-6ec7-4415-8fa9-e1342
Last Error	
Last Status	Complete
Last Run Time	3/25/19 8:47 PM
Run Duration	0 Hr. 0 Min. 46.01 Sec.
Grid Host	
User-Added Node	No

#### Number of terms=5024

Term	Role	Attribute	Status	Weight	Imported Frequency	Freq
elt	100	Alpha	Keep	0.113 0.027	238	238
inbroken		Alpha	Keep	0.027	12	12
nead	199911	Alpha	Keep	0.054	12 279	270
suddenly	100	Alpha	Keep	0.134	207	213
	2991			0.134	207 178	207
omina	197	Alpha	Keep	0.100	178	1/8
1985	100	Alpha	Keep	0.030	23	23
pegin	144	Alpha	Keep	0.100 0.030 0.131	23 67 79	67
sudden	100	Alpha	Keep	0.043	79	79
wanted	***	Alpha	Keep	0.123	112	112
millions	444	Alpha	Keep	0.123 0.063	29	29
nen	100	Alpha	Keep	0.197	440	440
ree	440	Alpha	Keep	0.197 0.193 0.143	440 153 66	153
drawn	100	Alpha	Keep	0.143	66	66
week		Alpha	Keep	0.102 0.202 0.139	49	49
evening		Alpha	Keep	0.202	118	118
wolc		Alpha	Keep	0.139	70	79
ums		Alpha	Keep	0.405	70	70
prepared	311	Alpha	Keep	0,400	79 70 48	12 279 207 178 23 677 112 29 440 153 66 49 118 770 488 633 558 1344
				0.139 0.221 0.154	88	40
net	144	Aloha	Keep	0.221	90	00
emain	1000	Alpha	Keep	U.154	. 63	_63
yay	100	Alpha	Keep	0.101	558	558
1ear	110	Alpha	Keep	0.088	134	134
effect	100	Alpha	Keep	0.195 0.095 0.168	107	107
unning	100	Aloha	Keep	0.095	116	116
00		Alpha	Keep	0.168	130	130 129 147
eave	1000	Aloha	Keep	0.062	129	129
un		Alpha	Keep	0.062 0.087	147	147
south	100	Alpha	Keep	0.257	108	108
/ears		Alpha	Keep	0.052	184	184
ourth	444	Alpha	Keep	0.257 0.052 0.117	60	108 184 60
oreak		Alpha	Keep	0.063	53	53
water	***	Alpha	Keep	0.448	1943	53 1943
eturning	300	Alpha	Keep	0.174	27	1343
stop		Alpha	Keep	0.226	27 118	27 118
sitting	100	Alpha		0.226 0.118	46	110
	744		Keep	0.110	-40	. 40
nundred	***	Alpha	Keep	0.092 0.063	141	46 141 39 47
airty	1991	Alpha	Keep	0.063	39	39
strike		Alpha	Keep	0.040	47	47
placed	244	Alpha	Keep	0.301	115	115
orce		Alpha	Keep	0.040 0.301 0.419	312	115 312
speak		Alpha	Keep	D.143		88
nand	100	Aloha	Keep	0.034 0.021 0.061	421 72 62	421
state	1000	Alpha	Keep	0.021	72	72
pread	-	Alpha	Keep	0.061	62	62
ood		Alpha	Keep	0.338 0.108	217	217
hort	***	Alpha	Keep	0.108	217 177	177
NI TOTAL	100	POUR .	1,000	0.100	17.7	111

d) Stop Words-No; POS-No; Stem-No

Node ID Imported Data Exported Data Exported Data Notes Train Variables  Parse Variable Language English Different Parts of Speech No Noun Groups Whith word Terms Find Entities Custom Entities Ilenore Ignore Parts of Speech Ignore Types of Entities Innore Types of Entities Innore Types of Attributes Synonyms Status Field Start List S	General	
Imported Data Exported Data Notes Train Variables  Parse Variable Language English Different Parts of Speech Noun Groups Ves Multi-word Terms Find Entities Custom Entities Ilenore Language Innore Parts of Speech Innore Parts of Speech Innore Types of Entities Innore Types of Entities Innore Types of Attributes Savnonyms Stem Terms Synonyms Stem Terms No Synonyms Stem Terms Synonyms Synonyms Stem Terms Synonyms Synonyms Stem Terms Synonyms Synonyms Stem Terms Synonyms Synonym		TextParsing4
Exported Data Notes Train Variables Jarse Parse Variable Parse Var		T EAAP GI SIITUP
Notes Train Variables  Jaise Parse Variable Language Lang		
Train Variables  3 Joise Parse Variable Parse Varia		
Variables  Janso Parse Variable Pars		
Parse Variable FILTERED Language Enallsh  Different Parts of Speech No Noun Groups Yes Multi-word Terms SASHELP,ENG MULTI None Custom Entities Illancic Language Filter Language Enallsh  Aux' 'Coni' 'Det' 'Interi' 'Part' 'Prinance Types of Entities Language Yes Language Filter Language		
Parse Variable FILTERED Language English  Different Parts of Speech No Noun Groups Yes Multi-word Terms SASHELP.ENG MULTI Find Entities None Custom Entities None Custom Entities Illenors Lanore Parts of Speech Lanore Types of Entities Lanore Types of Attributes Synonymis Stem Terms No Stem Terms No Stem Terms SASHELP.ENGSYNMS  Stem Terms No SASHELP.ENGSYNMS  Titles Stop List Slop Lis		
Language Different Parts of Speech Noun Groups Wes Noun Groups Wes Multi-word Terms SASHELP.ENG MULTI Find Entities Custom Entities Ilenore Ignore Parts of Speech Ignore Parts of Speech Ignore Types of Entities Ignore Types of Entities Ignore Types of Attributes Savnonyms Stem Terms No SASHELP.ENGSYNMS Iller Start List Stop List Select Languages Facet Languages Create Time Run ID Last Error Last Status Last Fror Last Status Last Run Time Run Duration Run Duration Run Duration Run Host		ETI TERED
Different Parts of Speech No Noun Groups Yes Multi-word Terms SASHELP.ENG MULTI = Find Entitles None Custom Entitles Illencore Innore Parts of Speech Innore Types of Entitles Innore Types of Entitles Innore Types of Entitles Innore Types of Entitles Innore Types of Attributes Synonyms Stem Terms No Synonyms Start List Stop List Stop List Select Languages Report Number of Terms to Display Status Create Time 3/25/19 8:39 PM Run ID 63123da7-b279-46eb-92ee-5b6d Last Error Last Status Complete Last Run Time 3/25/19 8:49 PM Run Duration O Hr. 0 Min. 44.95 Sec. Grid Host		
Different Parts of Speech		Eliulisi
Noun Groups Multi-word Terms Find Entities Fund Entities Custom Entities Innore Innore Parks of Speech Innore Types of Entities Innore Types of Entities Stem Terms Synonyms Start List Stop List Stop List Select Languages Report Number of Terms to Display Status Create Time Run ID Last Error Last Status Last Run Time Synonyms ASSHELP.ENGSYNMS JA25/19 8:39 PM 63123da7-b279-46eb-92ee-5b6d- Last Status Last Run Time AJ25/19 8:49 PM Run Duration Run Duration O Hr. 0 Min, 44.95 Sec. Grid Host		No
Multi-word Terms		
Find Entities		
Custom Entities		
Innoce		HOTE
Innore Parks of Speech		
Innore Types of Attributes		"Aux" 'Cont' 'Dot' 'Intort' 'Dart' 'Dr
Ignore Types of Attributes		Aux Colli Dec Iliceli Part Pin
Symonyms   Start List   Store   SASHELP, ENGSYNMS		'Num' 'Dunct'
Stem Terms		Hall Fallet
Synonyms		No
Start List		
Start List		SPOTED LENGSTHES
Stop List   Select Languages   Select Languages   Select Languages   Select Languages   20000   20000   Status   Status   Select Languages   Sel		
Select Languages		
Report   Number of Terms to Display   20000     Status   Create Time   3/25/19 8:39 PM   63123da7-b279-46eb-92ee-5b6d-		
Status   Create Time   3/25/19 8:39 PM		
Status   Create Time   3/25/19 8:39 PM	Number of Terms to Display	20000
Create Time         3/25/19 8:39 PM           Run ID         63123da7-b279-46eb-92ee-5b6d-           Last Error         Complete           Last Status         Complete           Last Run Time         3/25/19 8:49 PM           Run Duration         0 Hr. 0 Min. 44.95 Sec.           Grid Host         0 Hr. 0 Min. 44.95 Sec.		
Last Error  Last Status  Complete  Jast Run Time  3/25/19 8:49 PM  Run Duration  0 Hr. 0 Min. 44.95 Sec.  Grid Host		3/25/19 8:39 PM
Last Status Complete Last Run Time 3/25/19 8:49 PM Run Duration 0 Hr. 0 Min. 44.95 Sec. Grid Host	Run ID	63123da7-b279-46eb-92ee-5b6d-
Last Run Time 3/25/19 8:49 PM Run Durablon 0 Hr. 0 Min. 44,95 Sec. Grid Host		
Run Duration 0 Hr. 0 Min. 44.95 Sec. Grid Host	Last Status	Complete
Run Duration 0 Hr. 0 Min. 44.95 Sec. Grid Host	Last Run Time	3/25/19 8:49 PM
	Run Duration	0 Hr. 0 Min. 44.95 Sec.
User-Added Node No		
	User-Added Node	No

#### Number of terms=6280

Alpha   Keep   0.027   12   12   12   13   14   15   15   15   15   15   15   15	Term	Role	Attribute	Status	Weight	Imported Frequency	Freq
unbroken	felt		Alpha	Keen	0.113	238	238
Alpha   Keep   0.153   413   44   416		***			0.027		12
Micha   Mich		***				413	413
Mee   Alpha   Keep   0.030   23		***			0.103	270	276
Alpha   Keep   0.030   23					0.034	207	201
Alpha   Keep   0.030   23						170	470
Marted   Aloha   Keep   0.043   79		111				1/8	1/6
Marted   Aloha   Keep   0.043   79		***				23	23 67 87
Number   N		***			0.131	67	6/
Millions   Alpha   Keep   0.053   29			Alpha				87
Millions   Alpha   Keep   0.053   29		***			0.043	79	79
Millions   Alpha   Keep   0.053   29				Keep	0.123	112	79 112
Men	nillions		Alpha	Keep	0.063	29	29
Tee	nen		Alpha	Keep	0.197	440	440
Information   Alpha   Keep   0.143   66					0.193		153
week         Aloha         Keep         0.102         49           whering         Aloha         Keep         0.202         118         1           plow         Aloha         Keep         0.139         79         1           prepared         Aloha         Keep         0.405         70         1           prepared         Aloha         Keep         0.139         48         1         48         1         <		***			0.143	66	66
Seep		***			0.143	49	49
Alpha   Keep   0.139   79		***			0.102		118
sums         Aloha         Keep         0.405         70           prepared         Aloha         Keep         0.139         48           met         Aloha         Keep         0.221         88           remain         Aloha         Keep         0.154         63           say         Aloha         Keep         0.101         558         5           rear         Aloha         Keep         0.101         558         5           rear         Aloha         Keep         0.105         107         1           unning         Aloha         Keep         0.195         107         1           unning         Aloha         Keep         0.095         116         1           roo         Aloha         Keep         0.033         543         5           roo         Aloha         Keep         0.062         129         1           roo         Aloha         Keep         0.168         130         1           roo         Aloha         Keep         0.100         79         1           roo         Aloha         Keep         0.087         147         1           run		***			0.202		118
Alpha   Keep   0.139   48     Met   Alpha   Keep   0.221   88     Met   Alpha   Keep   0.154   63     Met   Alpha   Keep   0.154   63     Met   Alpha   Keep   0.101   558     Met   Alpha   Keep   0.101   558     Met   Alpha   Keep   0.101   558     Met   Alpha   Keep   0.088   134   1     Met   Met   Met   Met   Met     Met   Met   Met   Met   Met     Met   Met   Met   Met   Met     Met   Met   Met     Met   Met   Met   Met     Met   Met   Met		***			0.139	/9	79
net         Aloha         Keep         0.221         88           emain         Aloha         Keep         0.154         63           lay         Aloha         Keep         0.101         558         5           stear         Aloha         Keep         0.088         134         1           stear         Aloha         Keep         0.195         107         1           unning         Aloha         Keep         0.095         116         1           sp         O.000         Aloha         Keep         0.033         543         5           sp         O.000         Aloha         Keep         0.062         129         1           eave         Aloha         Keep         0.062         129         1         1           eave         Aloha         Keep         0.062         129         1		111			0.405	70	70
emain         Aloha         Keep         0.154         63           fay         Aloha         Keep         0.101         558         5           rear         Aloha         Keep         0.088         134         1           fffect         Aloha         Keep         0.195         107         1           unning         Aloha         Keep         0.095         116         1           oo         Aloha         Keep         0.033         543         5           op         Aloha         Keep         0.168         130         1           eave         Aloha         Keep         0.062         129         1           robably         Aloha         Keep         0.062         129         1           robably         Aloha         Keep         0.062         129         1           rurely         Aloha         Keep         0.081         23         1           rurely         Aloha         Keep         0.081         23         1           rurely         Aloha         Keep         0.093         356         3           rurely         Aloha         Keep         0.052         184<		111			0.139		48
day         Aloha         Keep         0.101         558         5           bear         Aloha         Keep         0.088         134         1         <	net	***			0.221	88	88
day         Aloha         Keep         0.101         558         5           bear         Aloha         Keep         0.088         134         1         <	emain		Alpha	Keep	0.154	63	63
Near   Aloha   Keep   0.088   134   156   166   176	sav		Alpha	Keep	0.101	558	558
### Alpha Keep 0.195 107 11							134
unning         Aloha         Keep         0.95         116         1           pop         Aloha         Keep         0.033         543         5           pop         Aloha         Keep         0.168         130         1           pop         Aloha         Keep         0.062         129         1           probably         Aloha         Keep         0.002         129         1           un         Aloha         Keep         0.087         147         1           surely         Aloha         Keep         0.081         23         3           many         Aloha         Keep         0.093         356         3           south         Aloha         Keep         0.257         108         1           rears         Aloha         Keep         0.052         184         1           ourth         Aloha         Keep         0.017         60         0           rears         Aloha         Keep         0.063         53         1           rall         Aloha         Keep         0.082         10         1           rall         Aloha         Keep         0.048		***					107
OD         Alpha         Keep         0.033         543         5           OD         Alpha         Keep         0.168         130         1           Bave         Alpha         Keep         0.062         129         1           Inobability         Alpha         Keep         0.087         147         1           Unity         Alpha         Keep         0.081         23         1           Inany         Alpha         Keep         0.081         23         1           Inany         Alpha         Keep         0.093         356         3           Inouth         Alpha         Keep         0.093         356         3           Inouth         Alpha         Keep         0.052         184         1           Inouth         Alpha         Keep         0.052         184         1           Inouth         Alpha         Keep         0.053         53           Inouth         Alpha         Keep         0.063         53           Inouth         Alpha         Keep         0.174         27           Inouth         Inouth         Keep         0.174         27 <t< td=""><td></td><td></td><td></td><td></td><td>0.006</td><td></td><td>116</td></t<>					0.006		116
Ope					0.033		543
eave Aloha Keep 0.062 129 1 probably Aloha Keep 0.100 79 un Aloha Keep 0.087 147 un Aloha Keep 0.087 147 un Aloha Keep 0.081 23 many Aloha Keep 0.081 23 many Aloha Keep 0.093 356 3 prouth Aloha Keep 0.052 184 1 prears Aloha Keep 0.052 184 1 prears Aloha Keep 0.053 53 preak Aloha Keep 0.063 53 reak Aloha Keep 0.063 53 rail Aloha Keep 0.082 108 1 prear Aloha Keep 0.448 1943 19 prunting Aloha Keep 0.448 1943 19 prunting Aloha Keep 0.226 118 1 prunting Aloha Keep 0.226 138 1 prunting Aloha Keep 0.092 141 1 prunting Aloha Keep 0.092 141 1 prunting Aloha Keep 0.093 39 prunting Aloha Keep 0.063 39 prunting Aloha Keep 0.040 47 placed Aloha Keep 0.040 47 placed Aloha Keep 0.040 47 proces Aloha Keep 0.419 312 33		***					343
run         Aloha         Keep         0.087         147         1 surely           Manny         Aloha         Keep         0.081         23           many         Aloha         Keep         0.093         356         3           south         Aloha         Keep         0.257         108         1           rears         Aloha         Keep         0.052         184         1           fourth         Aloha         Keep         0.117         60         0           break         Aloha         Keep         0.063         53         53         53         53         1 <td< td=""><td></td><td>***</td><td></td><td></td><td>0.168</td><td>130</td><td>130</td></td<>		***			0.168	130	130
un   Aloha   Keep   0.087   147   1   1   1   1   1   1   1   1   1			Alpha		0.062	129	129 79
surely         Aloha         Keep         0.081         23           many         Aloha         Keep         0.093         356         3           south         Aloha         Keep         0.257         108         1           rears         Aloha         Keep         0.052         184         1           ourth         Aloha         Keep         0.117         60           oreak         Aloha         Keep         0.083         53           rall         Aloha         Keep         0.082         108         1           vater         Aloha         Keep         0.448         1943         19           stop         Aloha         Keep         0.174         27         27           stop         Aloha         Keep         0.126         118         1           stop         Aloha         Keep         0.126         118         1           undered         Aloha         Keep         0.126         118         1           strike         Aloha         Keep         0.092         141         1           strike         Aloha         Keep         0.040         47           s					0.100		79
years         Aloha         Keep         0.052         184         1           fourth         Aloha         Keep         0.117         60           break         Aloha         Keep         0.083         53           sall         Aloha         Keep         0.082         108         1           water         Aloha         Keep         0.448         1943         19           eturning         Aloha         Keep         0.174         27         27           stop         Aloha         Keep         0.226         118         1           undred         Aloha         Keep         0.118         46           undred         Aloha         Keep         0.092         141         1           lairly         Aloha         Keep         0.063         39         strike           strike         Aloha         Keep         0.040         47           placed         Aloha         Keep         0.301         115         1           locce         Aloha         Keep         0.419         312         3	un			Keep	0.087		147
rears         Aloha         Keep         0.052         184         1           ourth         Aloha         Keep         0.117         60           break         Aloha         Keep         0.083         53           sall         Aloha         Keep         0.082         108         1           vater         Aloha         Keep         0.448         1943         19           sturning         Aloha         Keep         0.174         27         7           stop         Aloha         Keep         0.226         118         1           stifting         Aloha         Keep         0.118         46         1           sunder         Aloha         Keep         0.092         141         1           sirty         Aloha         Keep         0.063         39         1           strike         Aloha         Keep         0.040         47         1           slaced         Aloha         Keep         0.301         115         1           sloces         Aloha         Keep         0.419         312         3	surely			Keep	0.081		23
rears         Aloha         Keep         0.052         184         1           ourth         Aloha         Keep         0.117         60           break         Aloha         Keep         0.083         53           sall         Aloha         Keep         0.082         108         1           vater         Aloha         Keep         0.448         1943         19           sturning         Aloha         Keep         0.174         27         7           stop         Aloha         Keep         0.226         118         1           stifting         Aloha         Keep         0.118         46         1           sunder         Aloha         Keep         0.092         141         1           sirty         Aloha         Keep         0.063         39         1           strike         Aloha         Keep         0.040         47         1           slaced         Aloha         Keep         0.301         115         1           sloces         Aloha         Keep         0.419         312         3					0.093	356	356
rears         Aloha         Keep         0.052         184         1           ourth         Aloha         Keep         0.117         60           break         Aloha         Keep         0.083         53           sall         Aloha         Keep         0.082         108         1           vater         Aloha         Keep         0.448         1943         19           sturning         Aloha         Keep         0.174         27         7           stop         Aloha         Keep         0.226         118         1           stifting         Aloha         Keep         0.118         46         1           sunder         Aloha         Keep         0.092         141         1           sirty         Aloha         Keep         0.063         39         1           strike         Aloha         Keep         0.040         47         1           slaced         Aloha         Keep         0.301         115         1           sloces         Aloha         Keep         0.419         312         3					0.257	108	108
Ourth         Aloha         Keep         0.117         60           oreak         Aloha         Keep         0.063         53           rall         Aloha         Keep         0.082         108         1           vater         Aloha         Keep         0.448         1943         19           sturning         Aloha         Keep         0.174         27           stoo         Aloha         Keep         0.226         118         1           sitting         Aloha         Keep         0.118         46         1           sundred         Aloha         Keep         0.092         141         1           sitriy         Aloha         Keep         0.063         39         1           strike         Aloha         Keep         0.040         47         1           slaced         Aloha         Keep         0.301         115         1           sloce         Aloha         Keep         0.419         312         3		***					184
break         Aloha         Keep         0.063         53           sall         Aloha         Keep         0.082         108         1           water         Aloha         Keep         0.448         1943         19           eturning         Aloha         Keep         0.174         27         1           stop         Aloha         Keep         0.226         118         1           sitting         Aloha         Keep         0.118         46         1           undered         Aloha         Keep         0.092         141         1         1           airly         Aloha         Keep         0.063         39         1 </td <td></td> <td>***</td> <td></td> <td></td> <td></td> <td></td> <td>60</td>		***					60
all     Alpha     Keep     0.082     108     1       vater     Alpha     Keep     0.448     1943     19       eturning     Alpha     Keep     0.174     27       top     Alpha     Keep     0.226     118     1       titing     Alpha     Keep     0.118     46       bundred     Alpha     Keep     0.092     141     1       alify     Alpha     Keep     0.063     39       trike     Alpha     Keep     0.040     47       vlaced     Alpha     Keep     0.301     115     1       orce     Alpha     Keep     0.419     312     312		***					53
vater         Aloha         Keep         0.448         1943         19           eturning         Aloha         Keep         0.174         27           top         Aloha         Keep         0.226         118         1           litting         Aloha         Keep         0.18         46           undered         Aloha         Keep         0.092         141         1           airly         Aloha         Keep         0.063         39           tritike         Aloha         Keep         0.040         47           vlaced         Aloha         Keep         0.301         115         1           orce         Aloha         Keep         0.419         312         3					0.063		53
eturning Alpha Keep 0.174 27 top Alpha Keep 0.226 118 1 Ilting Alpha Keep 0.118 46 sundred Alpha Keep 0.092 141 1 airly Alpha Keep 0.063 39 trike Alpha Keep 0.040 47 slaced Alpha Keep 0.301 115 1 proce Alpha Keep 0.31 312 3		***			0.082		108
top         Aloha         Keep         0.226         118         1           itting         Aloha         Keep         0.118         46           sundred         Aloha         Keep         0.092         141         1           airly         Aloha         Keep         0.063         39           trirke         Aloha         Keep         0.040         47           slaced         Aloha         Keep         0.301         115         1           orce         Aloha         Keep         0.419         312         32		***			0.448	1943	1943
Itting Aloha Keep 0.118 46 Nundred Aloha Keep 0.092 141 1 Nundred Aloha Keep 0.063 39 Aloha Keep 0.040 47 Aloha Keep 0.040 47 Aloha Keep 0.301 115 1 Norse Aloha Keep 0.419 312 3		111					27
Itting Aloha Keep 0.118 46 Nundred Aloha Keep 0.092 141 1 Nundred Aloha Keep 0.063 39 Aloha Keep 0.040 47 Aloha Keep 0.040 47 Aloha Keep 0.301 115 1 Norse Aloha Keep 0.419 312 3	top	***	Alpha	Keep	0.226	118	118
Jundred          Aloha         Keep         0.092         141         1           Alipha         Keep         0.063         39         1         1trike          Aloha         Keep         0.040         47         47         47         48	itting		Alpha	Keep	0.118	46	46
airly Aloha Keep 0.063 39 Aloha Keep 0.040 47 Aloha Keep 0.301 115 1 Drice Aloha Keep 0.311 312 3							141
trike Ajoha Keep 0.040 47 Jaced Ajoha Keep 0.301 115 1 proce Ajoha Keep 0.419 312 33		***			0.063	30	39
Naced Alpha Keep 0.301 115 1		***			0.063		47
orce Alpha Keep 0.419 312 3		***			0.040		
orce Alpha Keep 0.419 312 3 Deak Alpha Keep 0.143 88 land Alpha Keep 0.034 421 4		***			0.301		115
peak Alpha Keep 0.143 88 land Alpha Keep 0.034 421 4		111					312
and Alpha Keep 0.034 421 4		***				88	88
tate Alpha Keen 0.021 72	and		Alpha	Keep	0.034	421	421
	state	***	Alpha	Keep	0.021	72	72

# Python Part

```
import string
import nltk
from nltk.tokenize import word tokenize
from nltk.stem import WordNetLemmatizer
from nltk.stem.snowball import SnowballStemmer
from nltk.corpus import wordnet as wn
from nltk.corpus import stopwords
from nltk.probability import FreqDist
nltk.download('punkt')
nltk.download('averaged perceptron tagger')
nltk.download('stopwords')
nltk.download('wordnet')
with open ("TextFiles/T1.txt", "r") as text_file:
adoc = text file.read()
# Convert to all lower case - required
a discussion = ("%s" %adoc).lower()
# Remove unwanted punctuation
a discussion = a discussion.replace
("''" != word) and ("``" != word) and (word!='description') and (word
!='dtype') \
and (word != 'object') and (word!="'s")]
print("\nDocument contains a total of", len(tokens), " terms.")
#POS Tagging
tagged tokens = nltk.pos tag(tokens)
pos_list = [word[1] for word in tagged_tokens if word[1] != ":" and \
word[1] != "."]
pos_dist = FreqDist(pos_list)
pos dist.plot(title="Parts of Speech")
for pos, frequency in pos dist.most common(pos dist.N()):
 print('{:<15s}:{:>4d}'.format(pos, frequency))
# Remove stop words
stop = stopwords.words('english') + list(string.punctuation)
stop tokens = [word for word in tagged tokens if word[0] not in stop]
# Remove single character words and simple punctuation
stop_tokens = [word for word in stop_tokens if len(word) > 1]
# Remove numbers and possive "'s"
stop tokens = [word for word in stop tokens \
if (not word[0].replace('.','',1).isnumeric()) and \
word[0]!="'s" ]
token_dist = FreqDist(stop_tokens)
print("\nCorpus contains", len(token_dist.items()), \
```

```
" unique terms after removing stop words.\n")
for word, frequency in token dist.most common(20):
 print('{:<15s}:{:>4d}'.format(word[0], frequency))
# Lemmatization - Stemming with POS
# WordNet Lematization Stems using POS
stemmer = SnowballStemmer("english")
wn_tags = {'N':wn.NOUN, 'J':wn.ADJ, 'V':wn.VERB, 'R':wn.ADV}
wnl = WordNetLemmatizer()
stemmed tokens = []
for token in stop_tokens:
term = token[0]
 pos = token[1]
 pos = pos[0]
 try:
 pos = wn tags[pos]
 stemmed_tokens.append(wnl.lemmatize(term, pos=pos))
 except:
 stemmed tokens.append(stemmer.stem(term))
#Token distribution
fdist = FreqDist(stemmed_tokens)
print("\nCorpus contains", len(fdist.items())," unique terms after
Stemming.\n")
# Top 20 terms
print('Top 20 terms sorted by frequency')
for word, freq in fdist.most_common(20):
 print('{:<15s}:{:>4d}'.format(word, freq))
fdist top = nltk.probability.FreqDist()
for word, freq in fdist.most_common(20):
 fdist_top[word] = freq
fdist_top.plot()
```

#### Document 1 – Case 1

Top 20 terms	sorted by frequen	cy Top 20 terms	sorted by frequency
say	: 360	say	: 360
tommy	: 300	tommy	: 300
one	: 291	one	: 291
would	: 269	would	: 269
man	: 248	man	: 248
could	: 220	could	: 220
go	: 212	go	: 212
come	: 211	come	: 211
eye	: 206	eye	: 206
make	: 284	make	: 204
see	: 179	see	: 179
men	: 177	men	: 177
like	: 173	like	: 173
get	: 171	get	: 171
upon	: 168	upon	: 168
know	: 165	know	: 165
look	: 155	look	: 155
professor	: 144	professor	: 144
	: 144		: 144
hand	: 142	hand	: 142

### Case 2

Corpus contains 6375 unique terms after Stemming.

```
Top 20 terms sorted by frequency
say
                  : 360
tommy
                  : 300
                  : 291
one
would
                 : 269
: 248
man
could
                  : 220
                  : 212
go
come
                  : 211
eye
                  : 286
make
                  : 284
                  : 179
: 177
: 173
: 171
see
men
like
get
upon
                  : 168
know
                  : 165
look
                  : 155
                  : 144
: 144
: 142
professor
hand
```

#### Case 3

```
Corpus contains 18273 unique terms after removing stop words.
               : 297
said
              : 269
: 253
would
one
              : 248
man
could
               : 220
men
               : 177
upon
               : 161
eyes
               : 155
like
               : 152
tomy
               : 151
n't
               : 144
               : 135
professor
               : 131
               : 120
came
two
               : 114
time
               : 114
holtz
               : 113
world
               : 187
               : 186
back
              : 100
```

#### Case 4

```
Top 20 words by frequency
the :5178
                  :4818
of
                  :2421
and
                  :2358
                  :1968
to
                  :1864
                  :1390
he
that
                  :1165
                  :1136
was
in
                  :1118
it
                  :1035
                  : 821
                  : 779
: 590
his
had
                  : 586
: 574
you
with
but
                  : 554
                  : 543
: 526
: 495
as
they
for
```

# Document 2 – 4 cases in order

Corpus contai	ns 3981 unique terms after Stem	ming.	ins 3981 unique terms after Stemming
Top 20 terms water make air light	sorted by frequency : 922 : 694 : 518 : 461	water make air light	sorted by frequency : 922 : 694 : 518 : 461
one	: 437	one	: 437
would	: 407	would	: 407
wire	: 334	wire	: 334
go	: 292	go	: 292
get	: 291	get	: 291
experiment	: 274	experiment	: 274
heat	: 258	heat	: 258
thing	: 241	thing	: 241
electricity	: 240	electricity	: 240
put	: 234	put	: 234
use	: 232	use	: 232
see	: 232	see	: 232
glass	: 222	glass	: 222
fig	: 214	fig	: 214
tube	: 212	tube	: 212
way	: 211	way	: 211

Corpus contain	ns 7368 uniq	ue terms		ontains a total of 108341 ds by frequency	terms.
	. 020		the	:8302	
water	: 920			:4935	
air	: 515		of	:3304	
would	: 407		a	:2772	
one	: 397		and	:2278	
light	: 289		to	: 2056	
wire	: 267		is	:2056	
electricity			it	: 2045	
glass	: 211		in	:1848	
fig	: 205		you	:1485	
way	: 204		that	:1115	
heat	: 198		CHOIC .	: 979	
illustration	: 198		water	: 920	
made	: 186			: 797	
make	: 180		on		
tube	: 168		as	: 745	
things	: 167		or	: 710	
electric	: 167		when	: 697	
see	: 162		not	: 666	
hot	: 157		are	: 647	
two	: 155		with	: 625	

#### Document 3 – all 4 cases

		Top 20 terms	sort	ed by	frequency
Corpus contai	ins 5282 unique terms after Stemming.	water	:	825	
		air		412	
	sorted by frequency	heat		388	
water	: 825	fig		365	
air	: 412	one		348	
heat	: 388				
fig	: 365	light		322	
one	: 348	use		263	
light	: 322	make		262	
use	: 263	illustration		258	
make	: 262	current		251	
illustration	: 258	force		218	
current	: 251				
force	: 218	gas		216	
gas	: 216	substance		197	
substance	: 197	would		195	
would	: 195	form		182	
form	: 182	color		181	
color	: 181	time		175	
time	: 175				
great	: 164	great		164	
sound	: 164	sound		164	
produce	: 164	produce	:	164	

```
Corpus contains 8865 unique Document contains a total of 104654 terms.
                              Top 20 words by frequency
                              the
                                               :8767
water
                : 816
                                               :4322
                              of
sin
                : 408
                                               :4228
fig
                : 349
                              and
                                               : 3248
                : 327
one
                                               :2719
                              a.
heat
                : 289
                              18
                                               :2421
illustration
                : 247
                                               :2286
                              in.
current
                : 232
                : 195
                              to
                                               :1941
would.
                              it.
                                               991
light
                : 188
                : 177
                                               : 900
gais
                              25
                                               : 874
force.
                : 167
                              by.
                : 157
                                               : 857
pressure
                              are
                : 156
                                               : 816
upon
                              water.
                : 152
                              that
                                               : 785
nay
                : 137
used
                              be.
                                               : 777
                : 132
bwo.
                                               : 722
                              which.
                : 129
motion
                                               : 633
                : 125
temperature
                                               : 581
                              or.
small.
                : 123
                              from
                                               : 551
made
                : 122
                              but:
                                               : 533
```

### Document 4

Ton 28 tons	s sorted by frequency	Top 20 terms sorted by frequency
catherine	: 485	catherine : 485
could	: 364	could : 364
say	: 327	say : 327
would	: 389	would : 309
go	: 239	go : 239
think	: 234	think : 234
know	: 223	know : 223
tilney	: 220	tilney : 220
one	: 211	one : 211
miss	: 207	miss : 207
must	: 190	must : 190
make	: 185	make : 185
well	: 184	well : 184
good	: 175	good : 175
room	: 172	room : 172
look	: 171	look : 171
much	: 170	much : 170
mrs.	: 168	mrs. : 168
time	: 167	time : 167
never	: 159	never : 159

Corpus conta		unique t	Document conta Top 20 words b	ins a total of 82698 y frequency
could	: 364		the	:3174
catherine	: 346			: 2793
would	: 309		of	:2358
tilney	: 201		and	:2304
one	: 191		to	:2239
must	: 190		her	:1562
said	: 180		a	:1536
never	: 159		i	:1282
well	: 159		in	:1265
time	: 149		;	:1172
room	: 143		was	:1112
might	: 138		it	:1105
general	: 138		she	:1097
isabella	: 136		not	:1041
miss	: 132		you	: 918
every	: 130		that	: 805
good	: 127		be	: 795
though	: 120		for	: 726
brother	: 119		had	: 703
thorpe	: 118		as	: 684

### Document 5 – all 4 cases

```
Corpus contains 5515 unique terms aft
                                         Corpus contains 5515 unique terms aft
Top 20 terms sorted by frequency
                                          Top 20 terms sorted by frequency
               : 518
spray
                                          spray
                                                        : 518
day
sail
                                         day
sail
                                                         : 337
               : 322
                                                         : 322
one
               : 312
                                                         : 312
                                          one
come
               : 276
                                          come
                                                         : 276
sea
               : 275
                                          sea
                                                         : 275
                : 275
island
                                          island
                                                         : 275
make
                : 237
                                          make
                                                         : 237
would
                : 222
                                          would
                                                         : 222
wind
               : 202
                                          wind
                                                         : 202
say
could
               : 201
                                          say
could
                                                         : 201
                : 195
                                                         : 195
ship
               : 185
                                          ship
                                                         : 185
good
               : 173
                                          good
                                                         : 173
sloop
               : 166
                                         sloop
time
                                                         : 166
time
                : 164
                                                         : 164
               : 157
voyage
                                                         : 157
                                          voyage
go
great
find
                : 154
                                                         : 154
                                          go
                                         great
find
                : 145
                                                         : 145
                : 144
                                                         : 144
```

Corpus	contains 8718	unique term		contains a total ords by frequency	of 75985	terms.
			the	:5833		
spray	: 517			: 2803		
one	: 281		of	:2370		
sea	: 226		and	:2121		
would	: 222					
day	: 216		a	: 2092		
island	: 205		i	:1893		
could	: 195		to	:1583		
came	: 167		in	:1370		
wind	: 161		was	:1335		
sloop	: 160		on	: 976		
			for	: 726		
time	: 152		that	: 722		
voyage	: 145					
good	: 144		it	: 703		
cape	: 128		at	: 643		
ship	: 127		had	: 550		
great	: 127		my	: 533		
days	: 121		with	: 527		
said	: 112		spray	: 518		
night	: 108		from	: 509		
many	: 108					
marry	. 100		as	: 488		

## Document 6- all cases

Corpus contains 3804 unique terms afte	Corpus contains 3804 unique terms after
Top 20 terms sorted by frequency	Top 20 terms sorted by frequency
time : 213	time : 213
come : 155	come : 155
one : 121	one : 121
upon : 113	upon : 113
say : 112	say : 112
little : 112	little : 112
go : 103	go : 103
thing : 100	thing : 100
could : 93	could : 93
	machine : 89
	saw : 89
saw : 89	
seem : 82	
look : 77	look : 77
hand : 77	hand : 77
like : 74	like : 74
see : 72	see : 72
think : 71	think : 71
man : 70	man : 70
make : 63	make : 63
find : 62	find : 62

Corpus contains 54	59 unique term		
time : 2	100	Top 20 words by	
		the	: 2241
	.14		:1763
little : 1	.12	i	:1265
upon : 1	.10	and	:1235
came : 1	.05	of	:1152
could :	93	a	: 815
said :	89	to	: 691
machine :	85	was	: 552
saw :	81	in	: 537
seemed :	71	my	: 437
man :	70	that	: 433
like :	69	it	: 418
thing :	66	had	: 355
traveller :	59	me	: 281
white :	59	as	: 261
would :	59	•	: 247
world :	52	at	: 238
still :	51	for	: 217
felt :	51	with	: 215
must :	49	time	: 200

# Document 7 – all 4 cases

Top 20	terms sorted by frequency	1	
tom	: 814	Top 20 terms	sorted by frequency
n't	: 627	tom	: 814
say	: 494	n't	: 627
go	: 374	say	: 494
get	: 316	go	: 374
would	: 287	get	: 316
boy	: 284	would	: 287
come	: 282	boy	: 284
huck	: 256	come	: 282
11	: 218	huck	: 256
time	: 216	11	: 218
know	: 203	time	: 216
one	: 202	know	: 203
could	: 202	one	: 202
well	: 191	could	: 202
take	: 178	well	: 191
see	: 172	take	: 178
make	: 169	see make	: 172
joe	: 166		: 169
	: 156	joe	: 166
tell	: 156	tell	: 156

Corpus contains 9426	unituue	contains a total of ords by frequency	79485
n't : 627		:3832	
tom : 521	the	: 3794	
said : 356	and	:3124	
would : 287	а	:1895	
'11 : 218	to	:1727	
could : 202	of	:1466	
	it	:1309	
time : 191	he	:1251	
one : 179			
huck : 168	was	:1180	
well : 158	that	:1022	
boys : 143	i	:1007	
joe : 142	in	: 955	
upon : 141	you	: 882	
little : 139	his	: 820	
tom : 138	tom	: 814	
got : 137	with	: 648	
never : 131	1	: 646	
boy : 121		: 643	
two : 120	n't	: 627	
came : 118	they	: 616	

## Document 8 – all 4 cases

Top 20 terms	sorted by frequency	Top 20 terms sorted by frequency	У
martian	: 247	martian : 247	
come	: 246	come : 246	
go	: 228	go : 228	
one	: 205	one : 205	
say	: 190	say : 190	
upon	: 172	upon : 172	
people	: 159	people : 159	
see	: 138	see : 138	
seem	: 134	seem : 134	
time	: 133	time : 133	
house	: 130	house : 130	
saw	: 129	saw : 129	
man	: 125	man : 125	
black	: 122	black : 122	
could	: 117	could : 117	
thing	: 115	thing : 115	
make	: 114	make : 114	
little	: 112	little : 112	
road	: 111	road : 111	
men	: 110	men : 110	

Corpus conta	ains 8296 unio	que termsDocument	contains a total of 6	4297
			rds by frequency	
one	: 184	the	:4794	
said	: 166		: 3004	
upon	: 164	and	: 2503	
martians	: 163	of	:2301	
people	: 159	a	:1635	
came	: 151	i	:1295	
man	: 125	to	:1175	
time	: 122	in	:1001	
black	: 122	was	: 854	
saw	: 118	that	: 793	
could	: 117	it	: 688	
men	: 110	had	: 582	
little	: 110	my	: 472	
		as	: 451	
road	: 104	with	: 449	
would	: 103	at	: 444	
brother	: 103	he	: 427	
us	: 102	on	: 378	
night	: 102	were	: 369	
way	: 100	for	: 348	
went	: 99	101	. 540	