In [1]:

```
import pandas as pd

data = pd.read_csv('Desktop/UCSD/Fall19/CSE250A/hw1_word_counts_05.txt', header = No
data.columns= [ 'Word', 'WordCount']
#print (type(data))
#print(data)
```

In [2]:

```
for i in range(6535):
    data['Prior_Probability']= data['WordCount']/sum(data['WordCount'])
```

In [3]:

```
data.sort_values(by= ['Prior_Probability'], axis=0,ascending= True,inplace= True )
```

In [4]:

```
#Sanity Check
data.tail(15)
```

Out[4]:

	Word	WordCount	Prior_Probability
5250	SIXTY	73086	0.009535
5806	THERE	86502	0.011286
6457	YEARS	88900	0.011598
2073	FORTY	94951	0.012388
4158	OTHER	106052	0.013836
1947	FIFTY	106869	0.013943
1975	FIRST	109957	0.014346
73	AFTER	110102	0.014365
6320	WHICH	142146	0.018545
5804	THEIR	145434	0.018974
18	ABOUT	157448	0.020542
6403	WOULD	159875	0.020858
1684	EIGHT	165764	0.021626
5102	SEVEN	178842	0.023333
5821	THREE	273077	0.035627

In [5]:

```
data.head(14)
```

Out[5]:

	Word	WordCount	Prior_Probability
3554	MAPCO	6	7.827935e-07
712	BOSAK	6	7.827935e-07
895	CAIXA	6	7.827935e-07
4160	OTTIS	6	7.827935e-07
5985	TROUP	6	7.827935e-07
1107	CLEFT	7	9.132590e-07
2041	FOAMY	7	9.132590e-07
977	CCAIR	7	9.132590e-07
5093	SERNA	7	9.132590e-07
6443	YALOM	7	9.132590e-07
5872	TOCOR	7	9.132590e-07
3978	NIAID	7	9.132590e-07
4266	PAXON	7	9.132590e-07
1842	FABRI	7	9.132590e-07

In [ ]:

```

```