1. The System was built based on the lectures. I added the DocumentVector class to represent the DocumentVector map talked about in the lecture. I used all the libraries already used in the provided code.

2.
When the threshold increases, the number of clusters increases but the size of each cluster decreases. I have included the number of clusters for each threshold but because the size of each cluster varies, I have not tabulated the size for each since it will be extremely large. In order to view each cluster, you can look at the .out files produced.

When thresh = .05
Number Of Clusters: 8

When thresh = .1
Number Of Clusters: 59

When thresh = .15
Number Of Clusters: 89

When thresh = .2
Number Of Clusters: 117

When thresh = .25
Number Of Clusters: 152

When thresh = .3
Number Of Clusters: 188

When thresh = .35
Number Of Clusters: 244

When thresh = .4
Number Of Clusters: 291

When thresh = .45
Number Of Clusters: 345

When thresh = .5
Number Of Clusters: 416

When thresh = .55
Number Of Clusters: 474

When thresh = .6

Number Of Clusters: 542

When thresh = .65
Number Of Clusters: 599

When thresh = .7
Number Of Clusters: 647

When thresh = .75
Number Of Clusters: 696

When thresh = .8
Number Of Clusters: 723

When thresh = .85
Number Of Clusters: 737

When thresh = .9
Number Of Clusters: 743

When thresh = .95
Number Of Clusters: 748

3.
From Single link, it uses the max similarity to calculate the score. It leads into long links that aren't good clusters because the end to end makes no sense.

From the complete link, it uses the min similarity to calculate the score. It creates good clusters but doesn't form many and leaves a ton of singletons.

Average link calculates the score using the average of all the similarities. It requires substantial similarity but it generally produces high quality clusters.

Mean link calculates the score using a centroid vector which is the average of all the document vectors in the cluster.

Overall, Mean link produces the most balanced results.