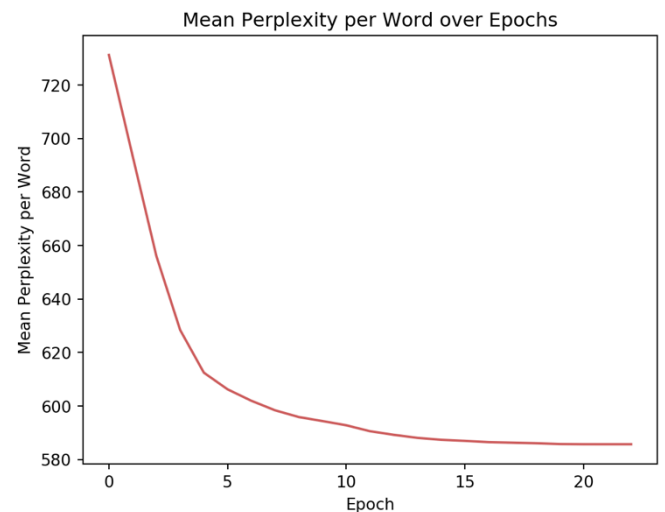
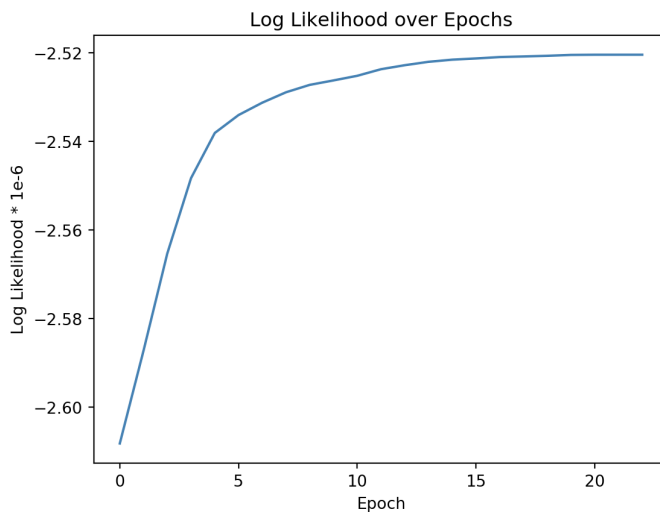


# Assignment3 – Report

Dorin Keshales – 313298424

Eran Hirsch – 302620745

1. The threshold chosen for stopping the EM iterations is 1.0.
2. The Log likelihood and the mean perplexity per word graphs:

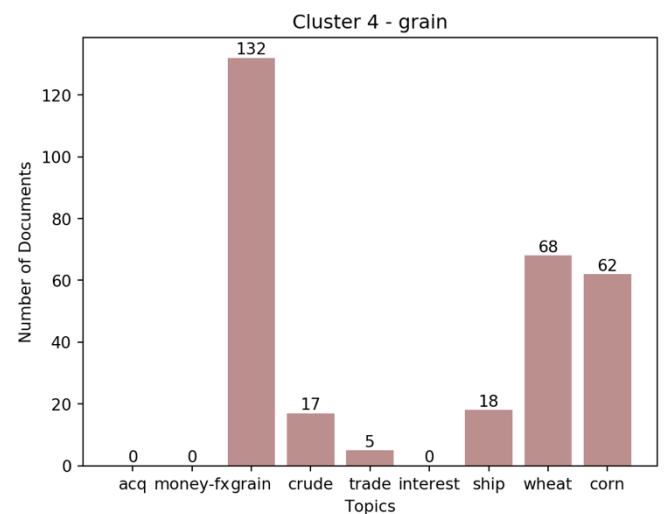
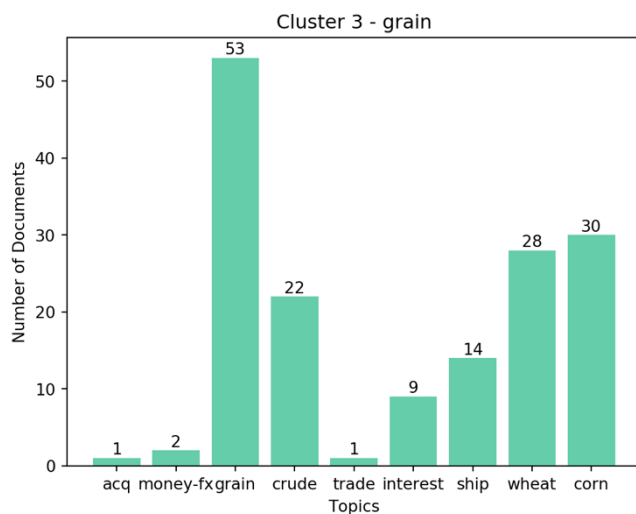
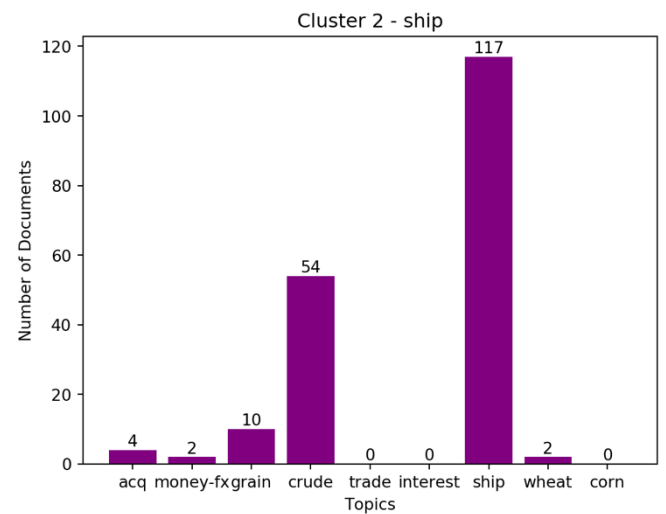
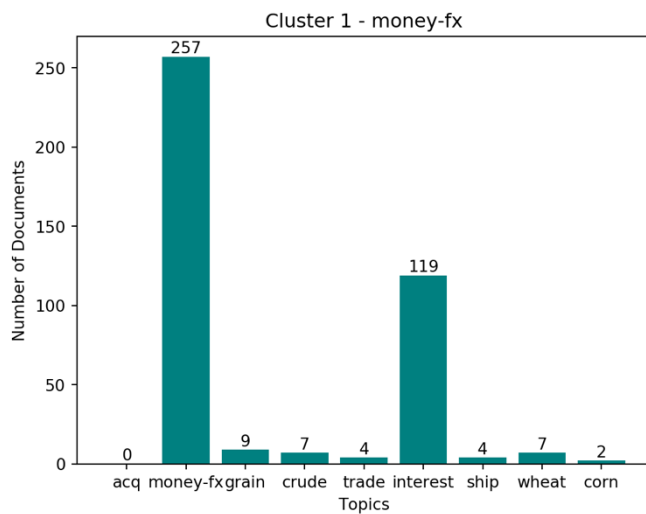


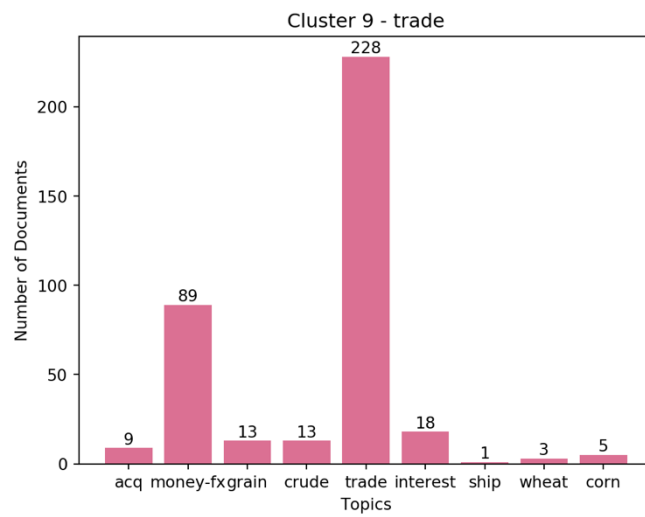
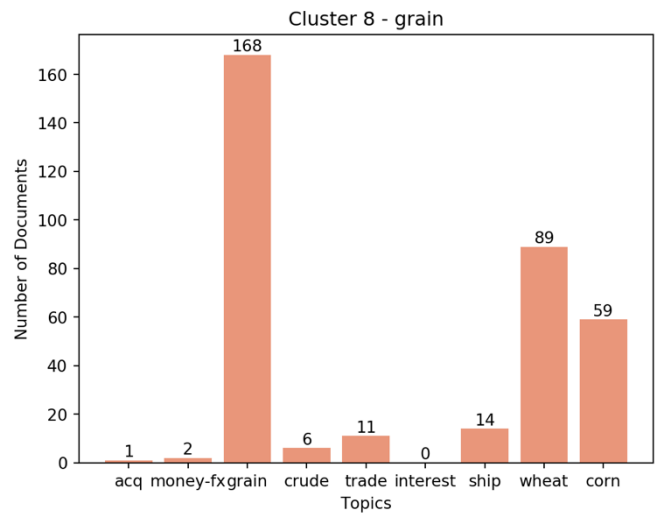
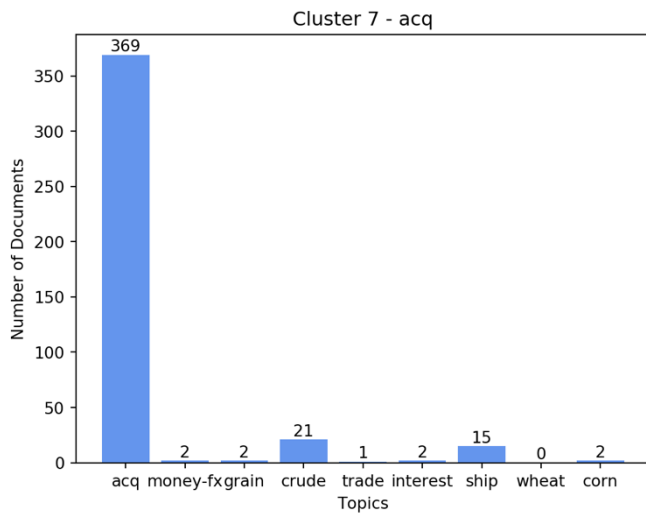
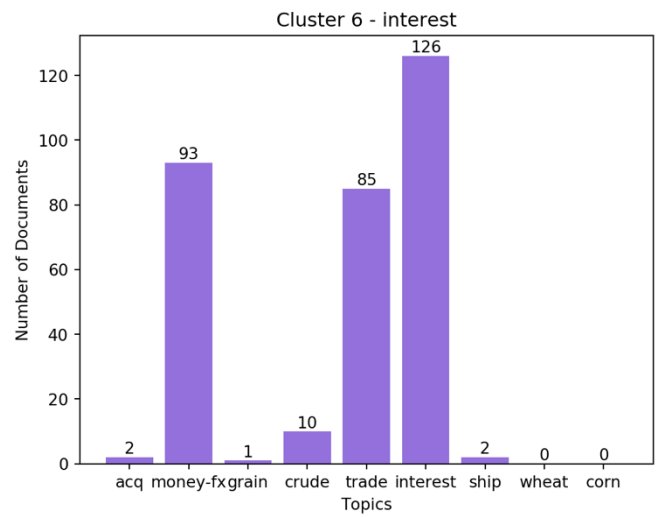
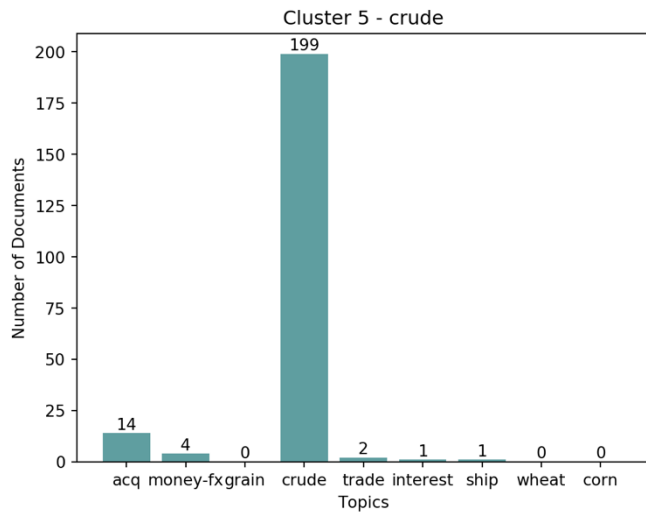
3. The confusion matrix:

Cluster ID	acq	money-fx	grain	crude	trade	interest	ship	wheat	corn	Cluster Size
7	369	2	2	21	1	2	15	0	2	400
9	9	89	13	13	228	18	1	3	5	334
1	0	257	9	7	4	119	4	7	2	312

6	2	93	1	10	85	126	2	0	0	262
5	14	4	0	199	2	1	1	0	0	211
8	1	2	168	6	11	0	14	89	59	192
4	0	0	132	17	5	0	18	68	62	163
2	4	2	10	54	0	0	117	2	0	152
3	1	2	53	22	1	9	14	28	30	98

#### 4. The 9 histograms of topics (one for each cluster):





5. Accuracy:  $0.776365 = 77.6365\%$

6. Epsilon = 0.0001

K = 10

Lambda = 1.0