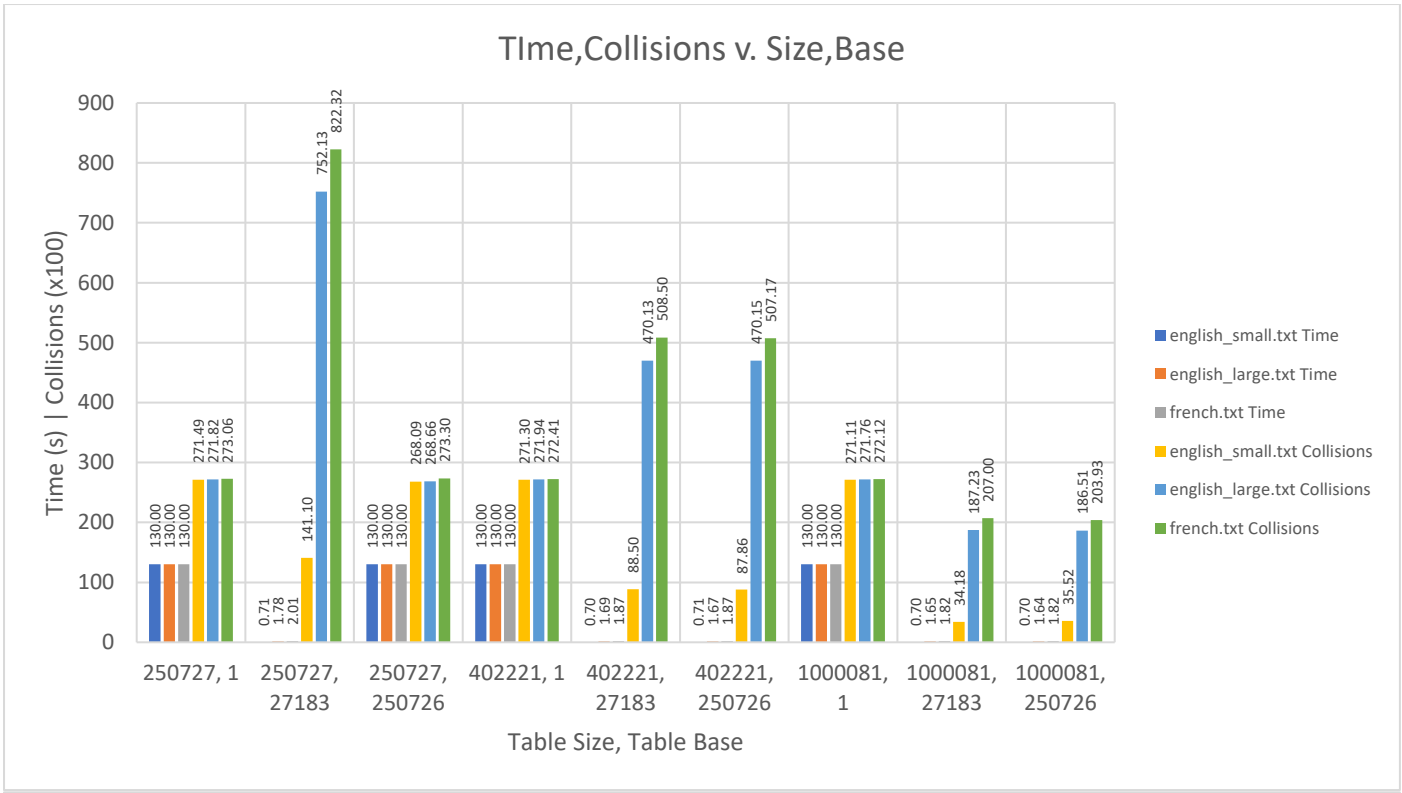
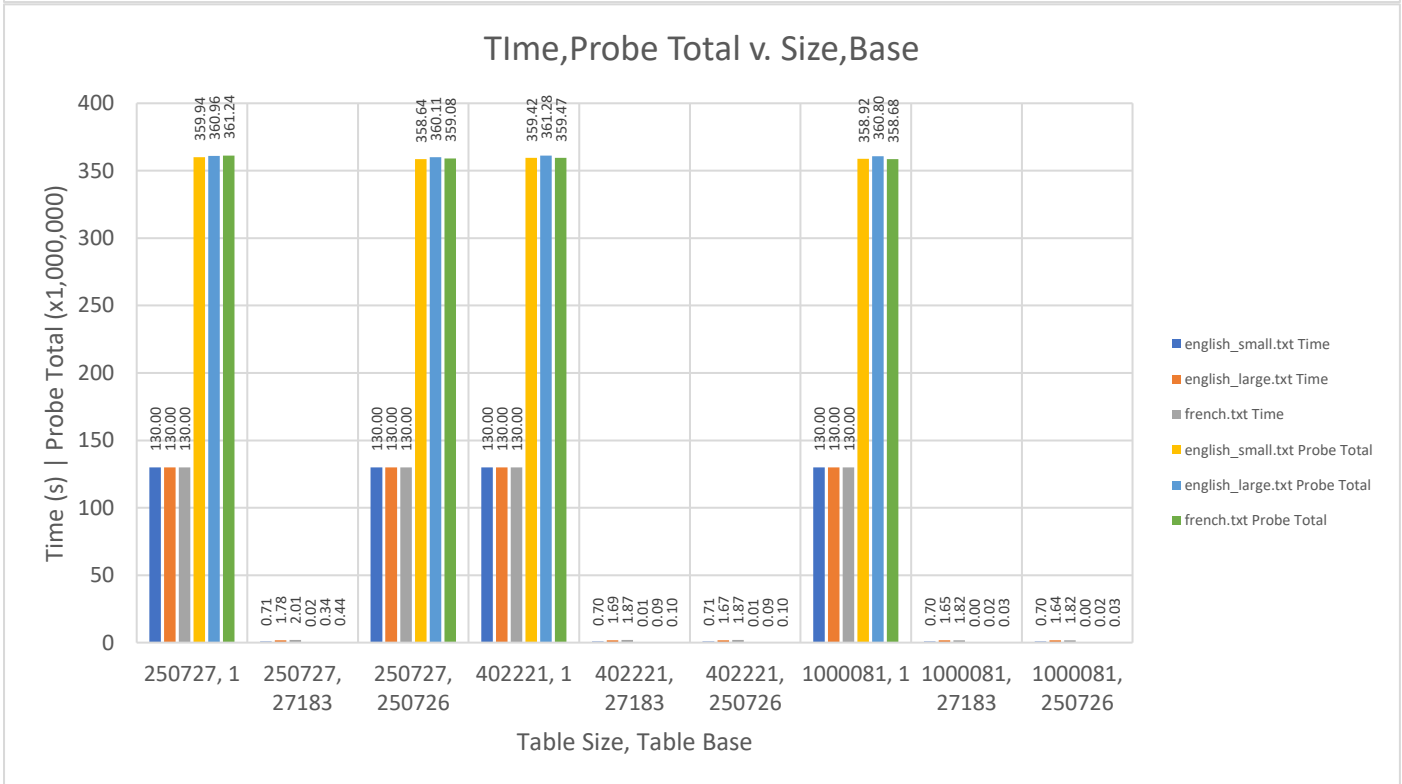
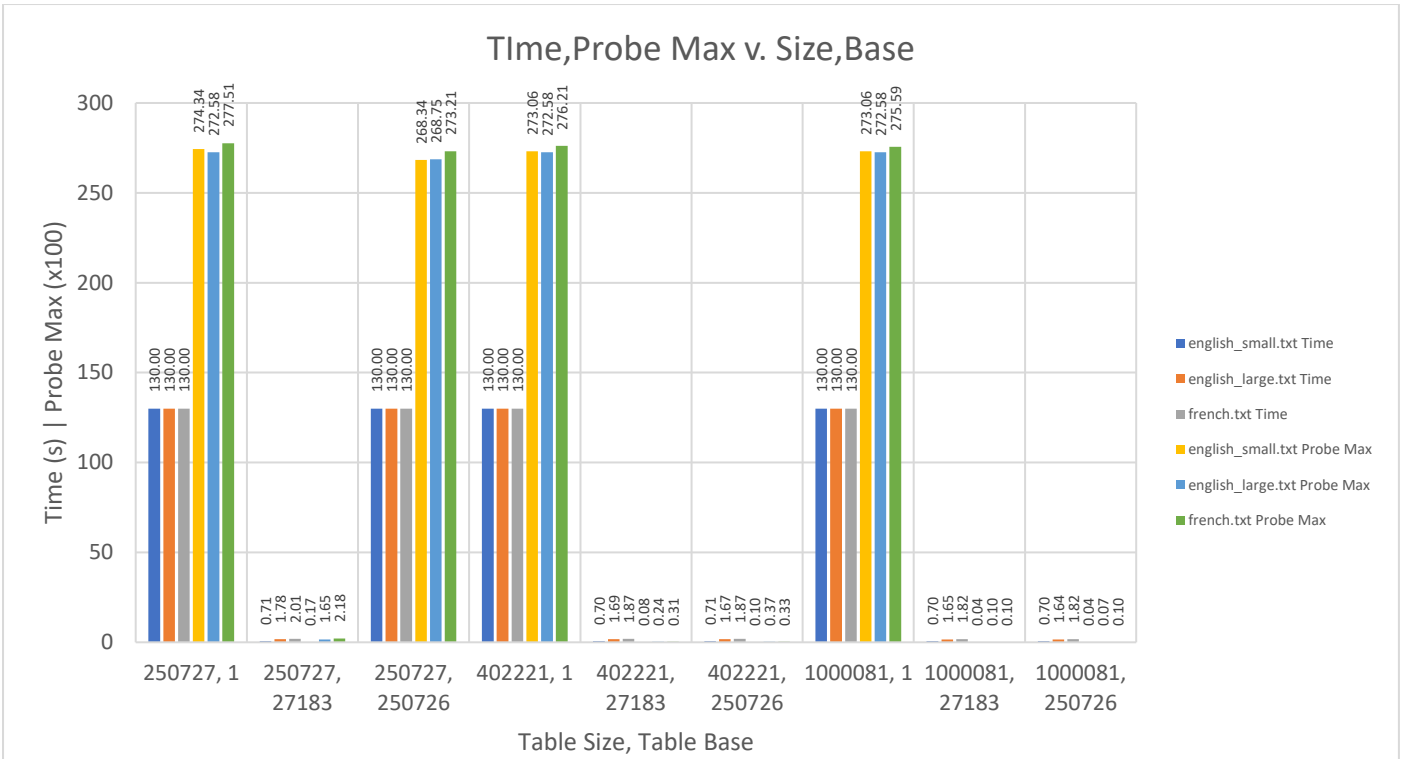
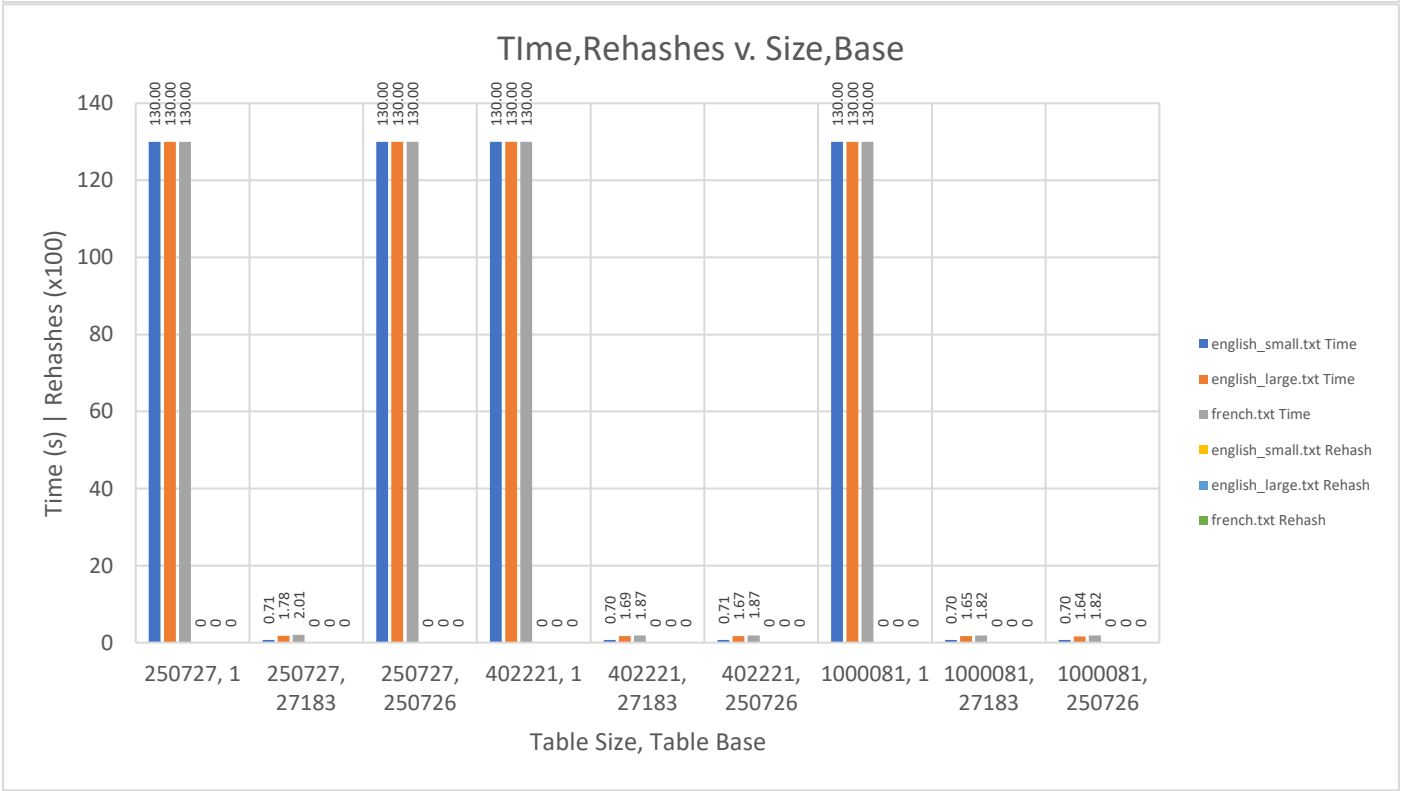

TIme,Collisions v. Size,Base
Time (s) | Collisions (x100)
900
800
700
600
500
400
300
200
100
0
130.00 130.00 130.00 271.49 271.82 273.06
0.71 1.78 2.01 141.10 752.13 822.32
130.00 130.00 130.00 268.09 268.66 273.30
130.00 130.00 130.00 271.30 271.94 272.41
0.70 1.69 1.87 88.50 470.13 508.50
0.71 1.67 1.87 87.86 470.15 507.17
130.00 130.00 130.00 271.11 271.76 272.12
0.70 1.65 1.82 34.18 187.23 207.00
0.70 1.64 1.82 35.52 186.51 203.93
250727, 1
250727, 27183
250727, 250726
402221, 1
402221, 27183
402221, 250726
1000081, 1
1000081, 27183
1000081, 250726
Table Size, Table Base
english_small.txt Time
english_large.txt Time
french.txt Time
english_small.txt Collisions
english_large.txt Collisions
french.txt Collisions


TIme,Probe Total v. Size,Base
Time (s) | Probe Total (x1,000,000)
400
350
300
250
200
150
100
50
0
130.00 130.00 130.00 359.94 360.96 361.24
0.71 1.78 2.01 0.02 0.34 0.44
130.00 130.00 130.00 358.64 360.11 359.08
130.00 130.00 130.00 359.42 361.28 359.47
0.70 1.69 1.87 0.01 0.09 0.10
0.71 1.67 1.87 0.01 0.09 0.10
130.00 130.00 130.00 358.92 360.80 358.68
0.70 1.65 1.82 0.00 0.02 0.03
0.70 1.64 1.82 0.00 0.02 0.03
250727, 1
250727, 27183
250727, 250726
402221, 1
402221, 27183
402221, 250726
1000081, 1
1000081, 27183
1000081, 250726
Table Size, Table Base
english_small.txt Time
english_large.txt Time
french.txt Time
english_small.txt Probe Total
english_large.txt Probe Total
french.txt Probe Total





- max_time = 120

- As shown in the graphs, a high collision count doesn't necessarily mean the execution time will be poor. It simply indicates that there is a possibility of long probe chains forming (clustering occurring within the hash table). It is this clustering that causes performance degradation.

- The better indicator of performance would be both the Probe Max statistic – which indicates the length of the longest probe – as well as the Probe Total statistic – total length probed. As the Hash Table uses linear probing, it can only probe in linear time and thus long probe chains degrade performance.

- All combinations executed here did not require a rehash as even the smallest hash table (size 250727) was larger than the maximum unique words observed (length 202358 – french.txt). The hash table will only rehash once it is completely full, and as such no rehashes occurred.

Overall, probe max and probe total are good indicators of cluster and performance degradation. They show that tables of larger sizes with appropriate bases are less susceptible to clustering and overall have better performance than their smaller counterparts.