Tanaka Jimha

MARKOV

1)

k = 1		
number of characters		Time(s)
100		0.104
200		0.132
400		0.224
800		0.393
1600		0.76

k = 5		
number of characters		Time(s)
100		0.032
200		0.063
400		0.123
800		0.256
1600		0.473

k =10		
number of characters		Time(s)
100		0.027
200		0.056
400		0.11
800		0.253
1600		0.471

Tanaka Jimha

The times decrease largely from k = 1 to k= 5, because the method scans the whole text each time and it has t do it for every character when k = 1. However between k = 5 and k = 10, the timings do not decrease substantially due to fewer number of scans for the 5 and 10 ngrams.

2) Since the brute force method scans the whole text for each ngram, I expect Hawthorne's scarlet to take longer than Romeo. I expect it to be (10/3) times longer than Romeo for each number of characters.

Hawthorne's	K =5	
Characters		Time
400		0.332
800		0.643
1600		1.286

No the empirical results do not match what I think because the times for Hawthorne's are smaller than expected given its size. This is because it is using the same k order markov model of 5 in both cases, and so the length of the scan itself does not have a significant impact on the time it will take.

I would expect the King James bible version to take roughly the same time as Hawthorne's and Romeo, without a very large difference.

3)

Map Model	k = 5	
Characters		time(s)
200		0.06
400		0.117
800		0.204
1600		0.374

The MapMarkovModel is significantly faster than the original brute force method. This is because the Map in which we store characters makes the process efficient, as we don't have to scan the text for each ngram.