Yiming Li A47222074

1. For each of the provided files, list the number of distinct words.

For alphabet.txt there are 26 distinct words.

For lines.txt there are 2268 distinct words.

For midsummer.txt there are 3015 distinct words.

For plurals.txt there are 4537 distinct words.

For scene1.txt there are 714 distinct words.

2. For the midsummer.txt file, identify which of your hash functions is best and justify your answer.

The hash algorithm A is best. Because the first one has lowest empty buckets percentage and lowest max number of items in a bucket, which means the how hashset is more well-distributed than the other two. So it will take less time when we used it to search in the future.

3. For the lines.txt file, identify which of your hash functions is best and justify your answer.

The hash algorithm A is best. Because the first one has lowest empty buckets percentage and lowest max number of items in a bucket, which means the how hashset is more well-distributed than the other two. So it will take less time when we used it to search in the future.

4. For the plurals.txt file, identify which of your hash functions is best and justify your answer.

The hash algorithm A is best. Because the first one has lowest empty buckets percentage and lowest max number of items in a bucket, which means the how hashset is more well-distributed than the other two. So it will take less time when we used it to search in the future.