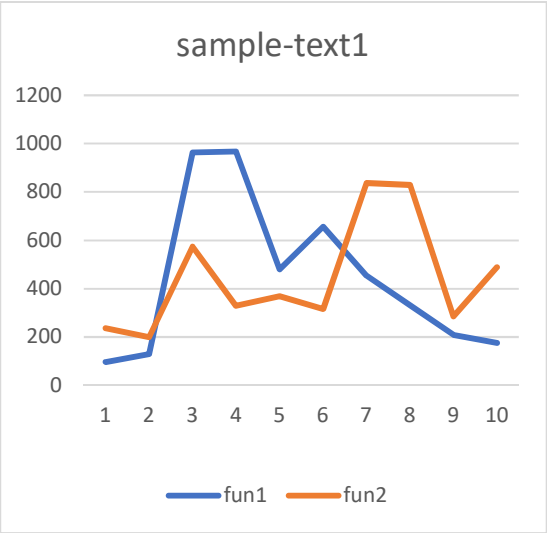
CO322: Data structures and algorithms

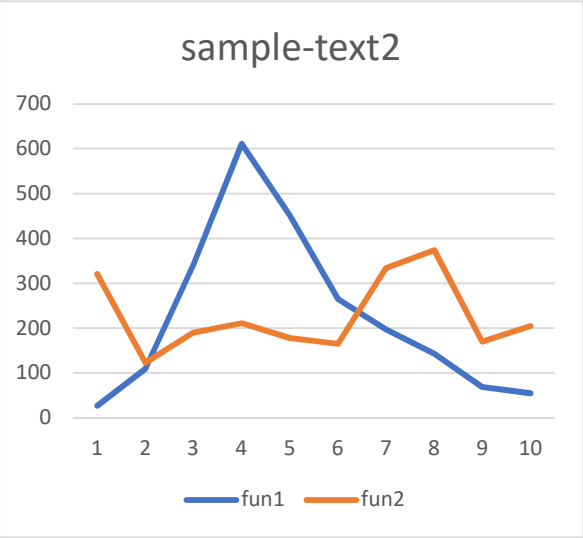
HASH TABLES

LAB-01

M.S.S.M PERERA
E/14/244
2-3-2018

WORDS DISTRIBUTION





	Hash Function 1		Hash Function 2	
	sample-text1	sample-text2	sample-text1	sample-text2
Bucket 1	96	27	236	321
Bucket 2	129	110	199	123
Bucket 3	963	342	573	190
Bucket 4	967	611	328	211
Bucket 5	480	452	369	178
Bucket 6	656	265	315	165
Bucket 7	454	197	837	334
Bucket 8	331	143	829	374
Bucket 9	209	69	285	170
Bucket 10	175	55	489	205

- Total words in sample-text1 : 4460
- Total words in sample-text2 : 2271

HASH FUNCTIONS

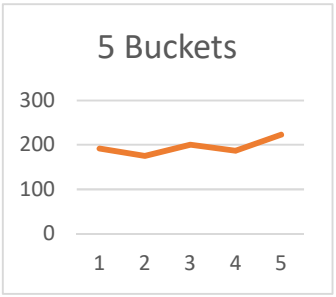
```
public int HashFun(String s,int buckets,int hashFunNo){

    int key = 0;
    if(hashFunNo==1){
        key = s.length() % buckets;
    }else if(hashFunNo == 2){
        int sum=0;
        char[] chars = s.toCharArray();
        int i=0;
        for(;i<chars.length;i++){
            sum += chars[i];
        }
    }

    return key
}
```

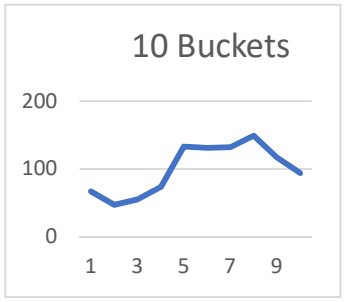
Hash function 1

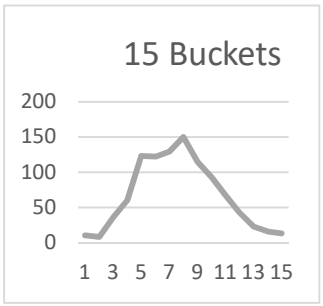
Hash function 2

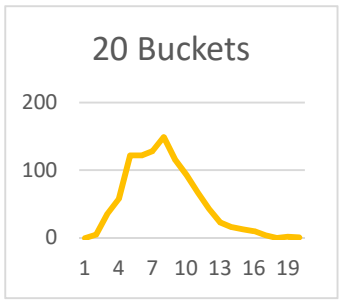
DIFFERENT NUMBER OF BUCKETS

Hash Function 1

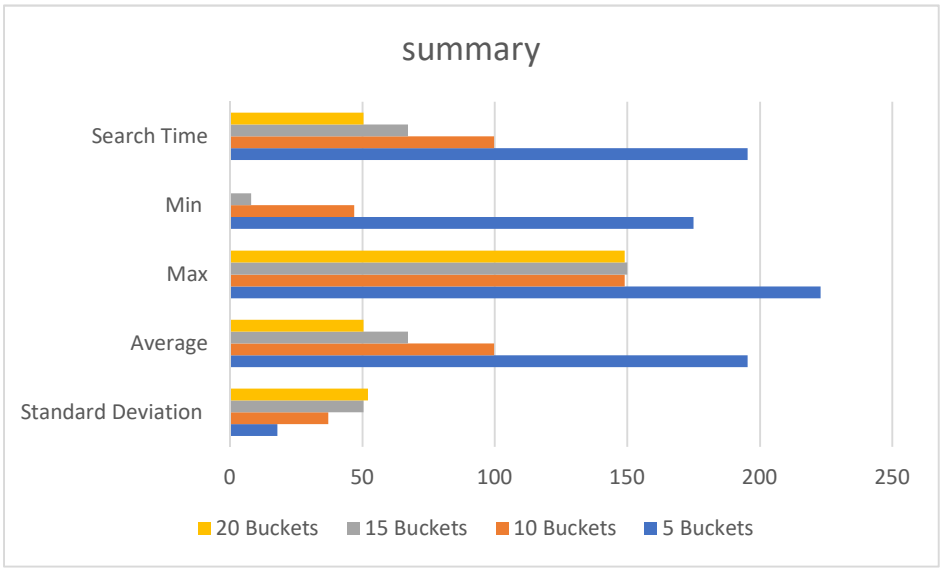






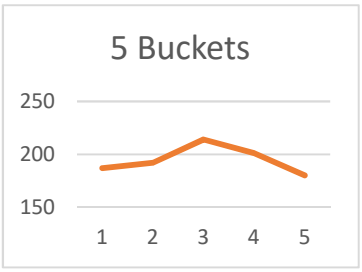


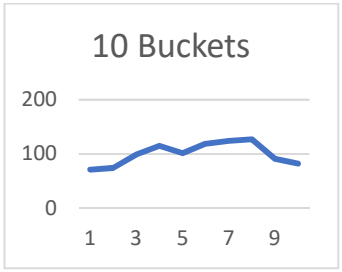
	5 Buckets	*10 Buckets*	*15 Buckets*	*20 Buckets*
Standard Deviation	17.89692711	37.19557919	50.45677075	52.12462293
Average	195.4	99.8	67.2	50.35
Max	223	149	150	149
Min	175	47	8	0
Search Time	195.4	99.8	67.2	50.35

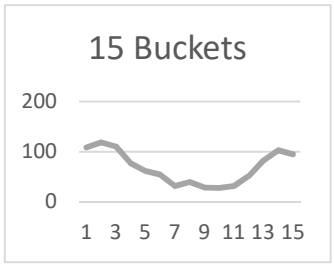


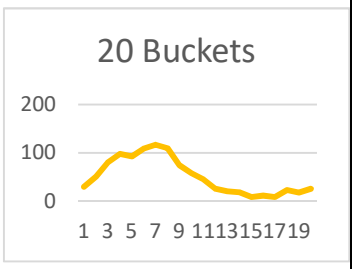
- Smallest searching time found in 20 bucket hash function but their standard deviation is more than 5 buckets, so 5 bucket hash method is good for the hashing.

Hash function 2

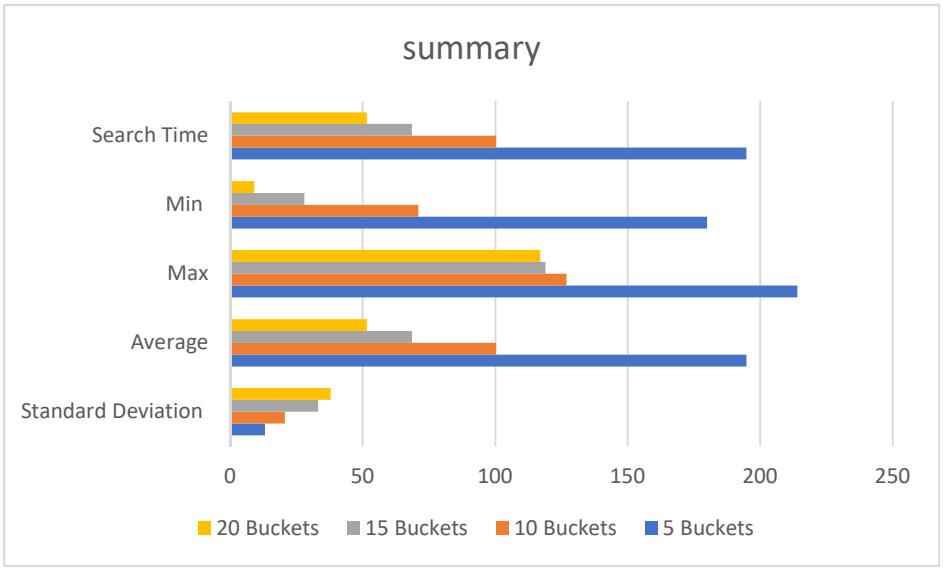






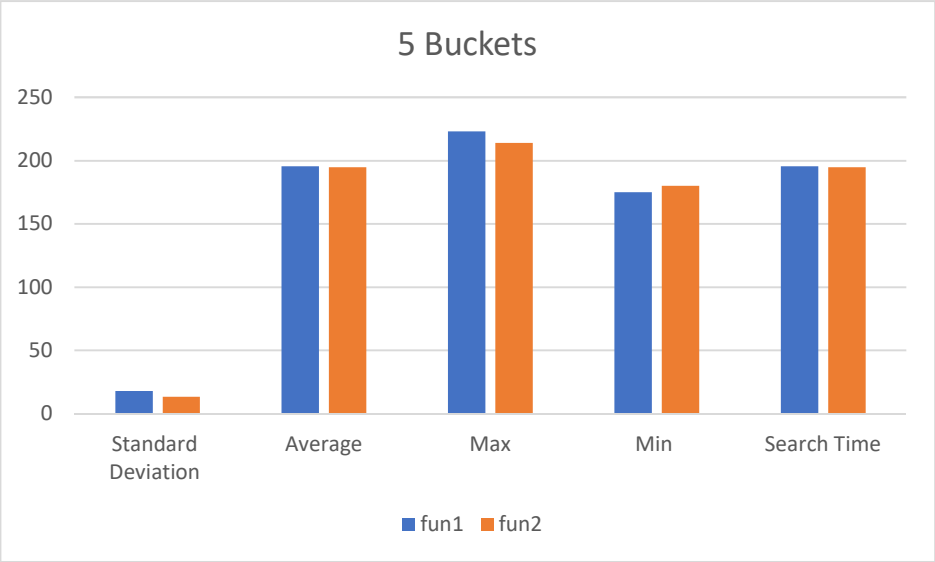


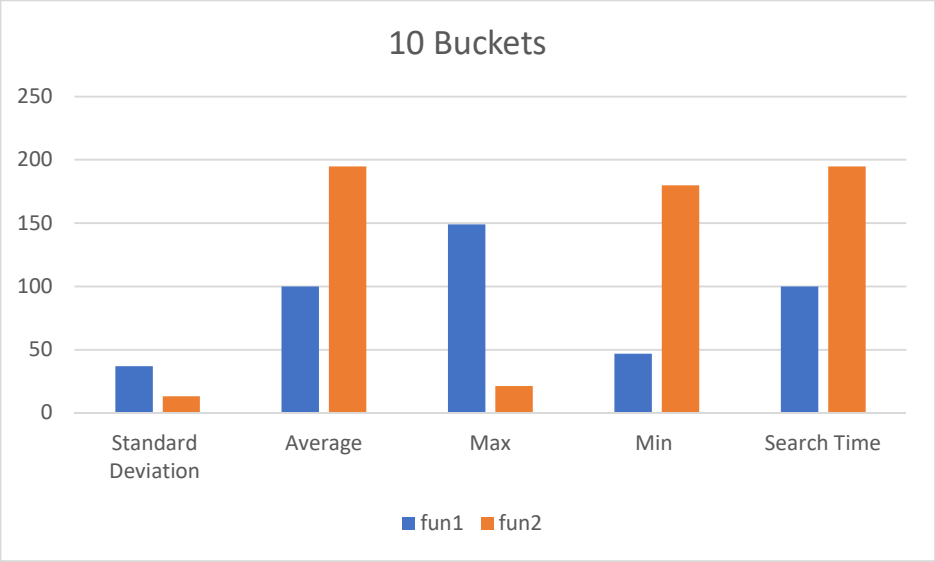
	5 Buckets	*10 Buckets*	*15 Buckets*	*20 Buckets*
Standard Deviation	13.17952958	20.58613341	33.10776056	37.85078043
Average	194.8	100.3	68.46666667	51.45
Max	214	127	119	117
Min	180	71	28	9
Search Time	194.8	100.3	68.46666667	51.45



- 2nd hash function's standard deviation is less than first hash function's standard deviation for all buckets.
- Considering above all hash functions, minimum standard deviation found in 5 bucket hash function. **So 2nd hash function is good for hashing.**

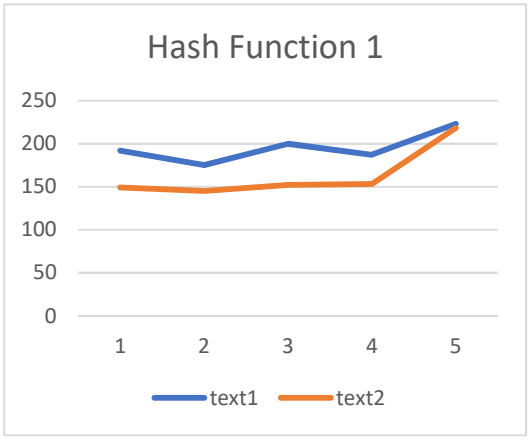
DIFFERENT NUMBER OF HASH FUNCTIONS

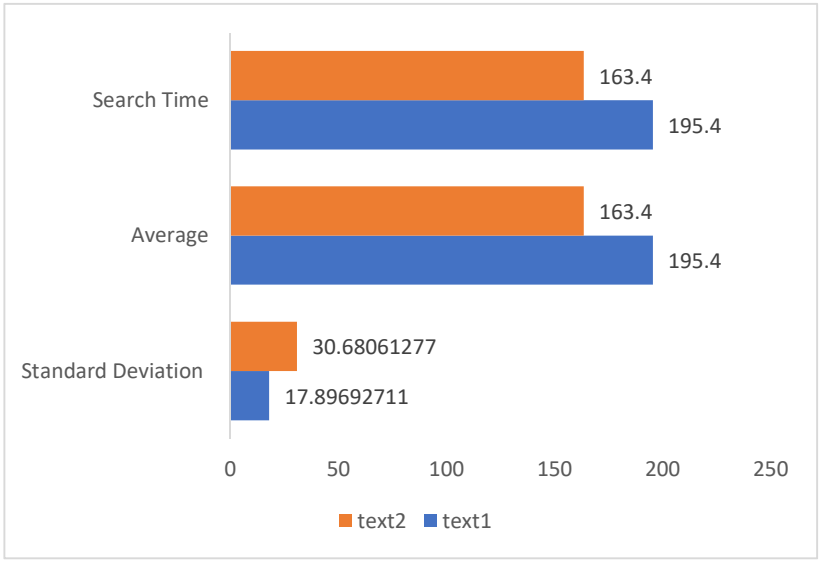


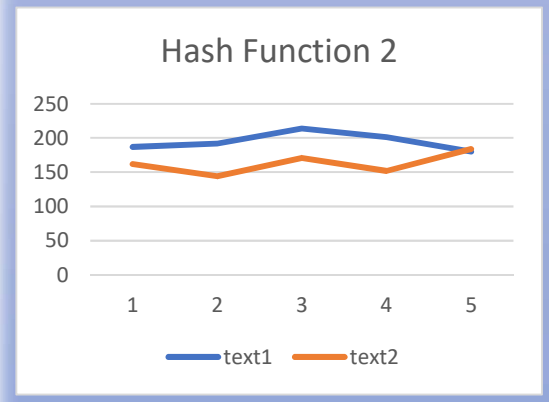


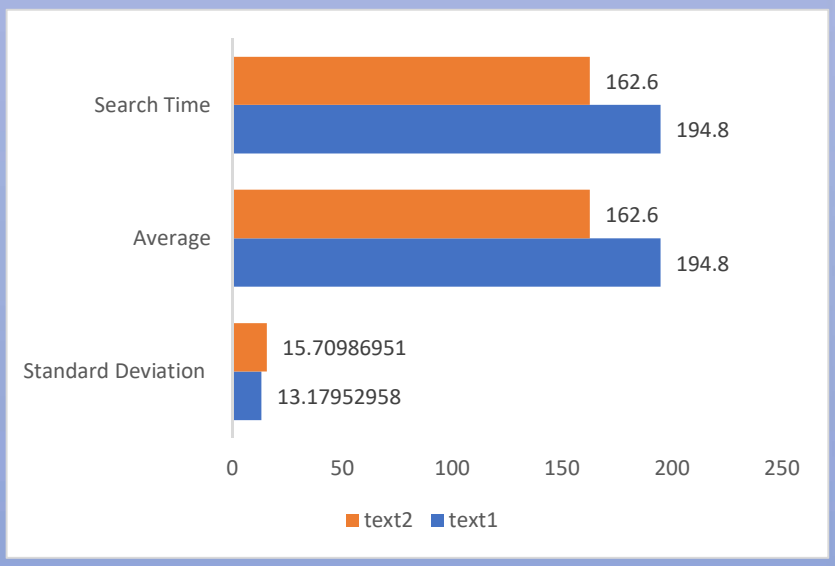
- 2nd hash function's searching time is more than 1st hash function's searching time.
- Hash function's standard deviation must be smallest to get perfect hash function performance.
- Considering above hash functions, 2nd function standard deviation is less than 1st function, **so 2nd hash function is the best hash function for these purposes.**

DIFFERENT TEXT FILES









- **This hash function is the best hash function, then try hash with deferent files (sample-text1 & sample-text2), So there is no big difference of distribution of the keys.**