Abdul Samad Gomda (ag7766)

HASH FUNCTIONS ANALYSIS OF PERFORMANCE

The document below evaluates the average number of collisions on four different hash functions considered below:

1. Bitwise/ Cyclic Shift Hash Function
2. Summation Hash Function
3. Polynomial Hash Function
4. XOR Hash Function

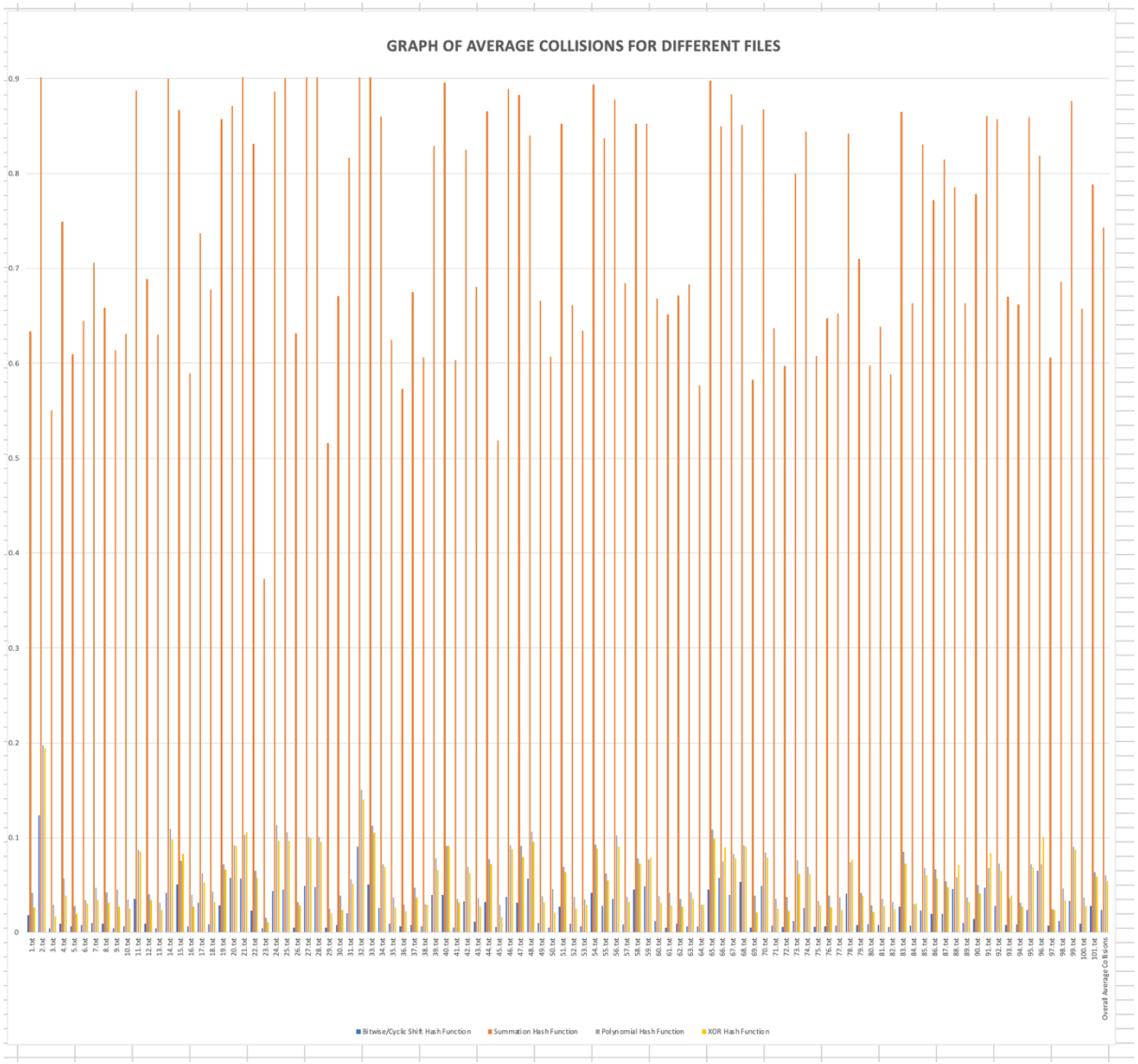
The average number of collisions was calculated in the program by dividing the total number of collisions by the size of the hash table, and recorded in the table below.

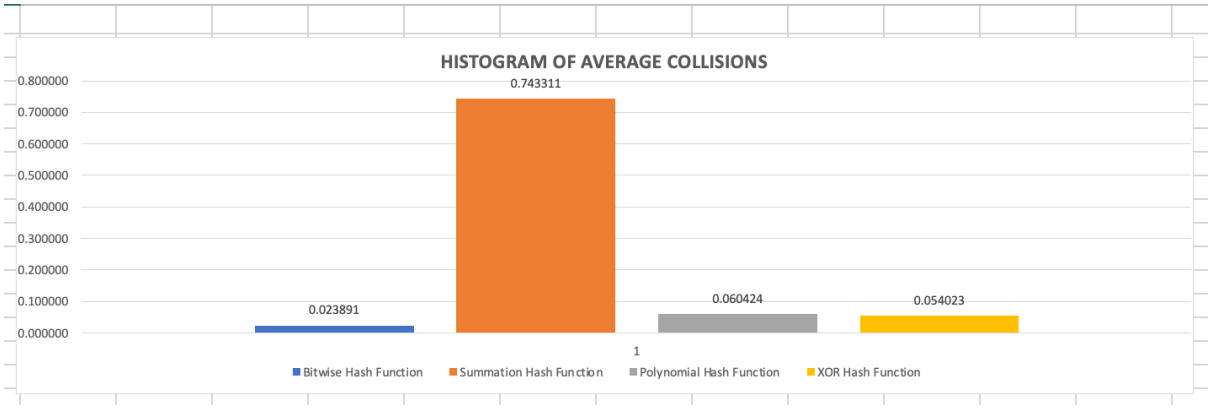
File Name	Bitwise/Cyclic Shift Hash Function	Summation Hash Function	Polynomial Hash Function	XOR Hash Function
1.txt	0.018556	0.6334	0.041625	0.02658
2.txt	0.123514	0.956757	0.197466	0.194477
3.txt	0.004274	0.550427	0.02906	0.017094
4.txt	0.009293	0.749495	0.05697	0.039192
5.txt	0.006489	0.609345	0.027904	0.019468
6.txt	0.00831	0.644875	0.034349	0.030471
7.txt	0.010242	0.705773	0.047486	0.033985
8.txt	0.009366	0.658402	0.042424	0.031405
9.txt	0.00436	0.613372	0.045058	0.027616
10.txt	0.006365	0.630808	0.03501	0.025461
11.txt	0.035855	0.88779	0.087433	0.08517
12.txt	0.009323	0.688911	0.040236	0.033857
13.txt	0.004572	0.630307	0.031352	0.024167
14.txt	0.041465	0.900286	0.109093	0.097937
15.txt	0.050452	0.867192	0.075726	0.082837
16.txt	0.006627	0.589838	0.039764	0.027246
17.txt	0.031585	0.737187	0.062277	0.052443
18.txt	0.008722	0.677783	0.043099	0.032324
19.txt	0.02885	0.8571	0.071826	0.066566
20.txt	0.057635	0.871173	0.092019	0.091032
21.txt	0.056614	0.910334	0.103337	0.106072
22.txt	0.023427	0.831447	0.065118	0.057772
23.txt	0.004644	0.373065	0.01548	0.010836
24.txt	0.044006	0.885858	0.11311	0.097066
25.txt	0.045151	0.900916	0.106052	0.096506
26.txt	0.005453	0.63122	0.032038	0.02863
27.txt	0.049467	0.908338	0.101049	0.099526
28.txt	0.047764	0.904473	0.100957	0.095619
29.txt	0.00555	0.516189	0.024977	0.020352
30.txt	0.00776	0.67046	0.039317	0.023797

31.txt	0.020764	0.81686	0.056063	0.05108
32.txt	0.090449	0.911086	0.150199	0.139852
33.txt	0.050687	0.905716	0.112539	0.104741
34.txt	0.025689	0.859806	0.071742	0.069079
35.txt	0.009524	0.624762	0.036825	0.026667
36.txt	0.006395	0.573142	0.029576	0.022382
37.txt	0.007748	0.675152	0.047039	0.037078
38.txt	0.006739	0.605795	0.030323	0.02965
39.txt	0.039557	0.829077	0.078524	0.06583
40.txt	0.039851	0.895623	0.091648	0.09106
41.txt	0.005579	0.603208	0.035565	0.031381
42.txt	0.032604	0.825012	0.069302	0.063317
43.txt	0.011302	0.680098	0.036364	0.027027
44.txt	0.032012	0.865345	0.07768	0.072247
45.txt	0.005731	0.518625	0.029608	0.016237
46.txt	0.037824	0.889107	0.092349	0.087683
47.txt	0.031472	0.882852	0.091045	0.079805
48.txt	0.056893	0.840123	0.106379	0.095165
49.txt	0.010005	0.665833	0.038019	0.031516
50.txt	0.005389	0.60662	0.046189	0.020785
51.txt	0.027476	0.852818	0.069303	0.064053
52.txt	0.009195	0.661494	0.037356	0.025287
53.txt	0.006944	0.63447	0.034722	0.029672
54.txt	0.041872	0.893576	0.092775	0.088567
55.txt	0.027671	0.837653	0.062347	0.054991
56.txt	0.035866	0.877951	0.102309	0.090311
57.txt	0.008392	0.684848	0.037762	0.032168
58.txt	0.044995	0.852755	0.078236	0.072548
59.txt	0.048635	0.85247	0.077015	0.079181
60.txt	0.01195	0.667782	0.038241	0.031549
61.txt	0.005099	0.65137	0.042065	0.028681

62.txt	0.009474	0.671579	0.035263	0.027368
63.txt	0.0064	0.6832	0.042667	0.035733
64.txt	0.006731	0.576664	0.02917	0.02917
65.txt	0.045056	0.898063	0.108359	0.098573
66.txt	0.05729	0.849542	0.074779	0.090144
67.txt	0.039394	0.883566	0.083217	0.078322
68.txt	0.053616	0.850882	0.092011	0.090363
69.txt	0.005464	0.582358	0.039032	0.021077
70.txt	0.04909	0.867741	0.084154	0.079109
71.txt	0.007033	0.636829	0.035806	0.024936
72.txt	0.005827	0.597232	0.037873	0.023307
73.txt	0.012458	0.800061	0.075965	0.061683
74.txt	0.025637	0.844129	0.06939	0.062041
75.txt	0.005926	0.607407	0.033333	0.028889
76.txt	0.006981	0.647469	0.038976	0.02676
77.txt	0.007365	0.652125	0.036827	0.023796
78.txt	0.040802	0.841839	0.074099	0.076556
79.txt	0.008377	0.710317	0.041446	0.038801
80.txt	0.00853	0.597769	0.028871	0.021654
81.txt	0.007995	0.638376	0.03567	0.02829
82.txt	0.005917	0.588018	0.031805	0.024408
83.txt	0.027522	0.864921	0.085242	0.072746
84.txt	0.007261	0.663382	0.030083	0.030602
85.txt	0.023004	0.830401	0.06811	0.060217
86.txt	0.019487	0.772077	0.066354	0.057227
87.txt	0.019715	0.814677	0.053888	0.048193
88.txt	0.045616	0.785664	0.058224	0.071115
89.txt	0.010194	0.663107	0.036893	0.032039
90.txt	0.013974	0.777875	0.049782	0.040757
91.txt	0.047423	0.860817	0.067894	0.083942
92.txt	0.028115	0.85724	0.073018	0.064965
93.txt	0.007862	0.669811	0.036164	0.039308
94.txt	0.008488	0.661538	0.0313	0.027586
95.txt	0.024061	0.859698	0.071858	0.068769
96.txt	0.064964	0.819062	0.07236	0.100724
97.txt	0.007427	0.606347	0.024308	0.023633
98.txt	0.011856	0.686275	0.046512	0.0342

97.txt	0.007427	0.606347	0.024308	0.023633
98.txt	0.011856	0.686275	0.046512	0.0342
99.txt	0.033539	0.876496	0.090622	0.087202
100.txt	0.009086	0.657581	0.036911	0.027825
101.txt	0.028076	0.78877	0.063878	0.059167
Overall Average Collisions	**0.023891416**	**0.743311436**	**0.060424109**	**0.054022683**





Conclusion of Analysis

From analysis the data set on the 101 text files, the hash function producing the least number of collisions is the Bitwise / Cyclic Shift Hash function, producing about 23 collisions per 1000 entries. Hence it has been set as the default hash function for the word count wizard.

The XOR hash Function follows in second place producing about 54 collisions per 1000 entries.

The Polynomial hash function follows in a close third place producing about 60 collisions per 1000 entries.

The worst of all four hash functions is the summation hash function which produces in upwards of 700 collisions per 1000 entries.