Part 2 – Results

The five documents were compressed using the Huffman encoding algorithm with three different canonical collections. The results after compression are shown in Table 1.

Table 1 - Size of the documents after compression with Huffman encoding with different canonical collections

		Encoded		
	Original [KB]	**Collection 1 [KB]**	**Collection 2 [KB]**	**Collection 3 [KB]**
EarthASCII	438	687.5	250	250
MysteryASCII	445	737.5	262.5	262.5
MythsASCII	741	1325	462.5	450
SimakASCII	310	500	175	175
WodehouseASCII	403	650	237.5	225
Total	2337	3900	1387.5	1362.5

From Table 1, it is evident that the choice of collection used affects the efficiency of the algorithm. Collection 3 produces the best result and this is justified as it uses a small number of large documents. This allows for more coverage of all characters and hence a shorter bit pattern in a code string. Collection 1 produces the worst result as it just contains a list of words and provides the least coverage.