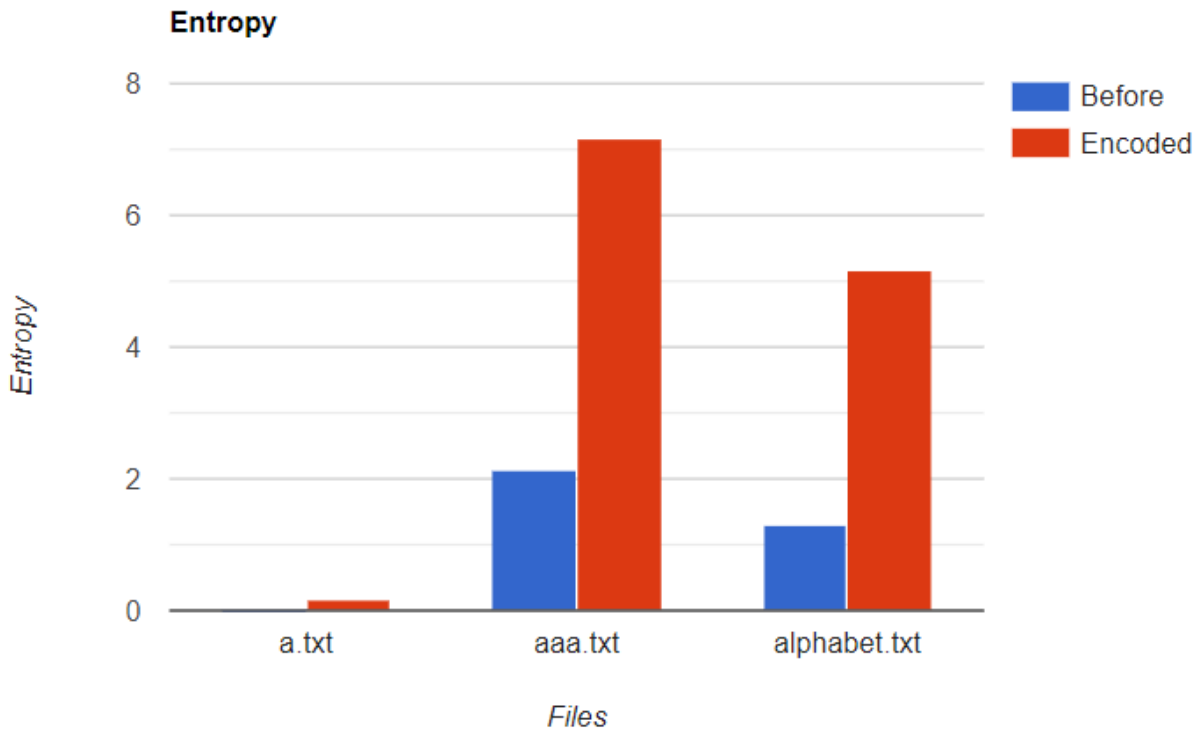
Devan O'Boyle
Assignment 5: Hamming Codes

Write Up

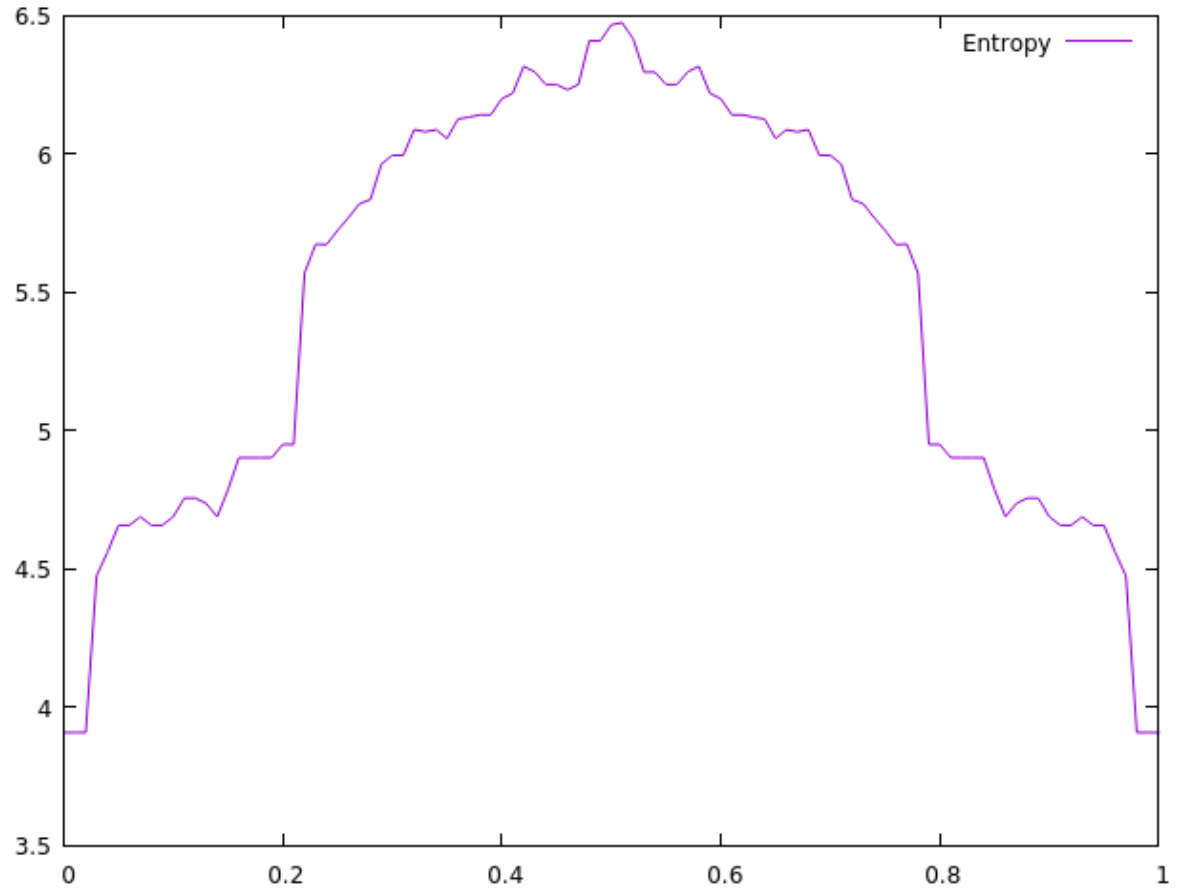
Entropy is defined as the amount of uncertainty or randomness that is collected by a program or operating system.

In this case, we were encoding and decoding a series of messages. Before we encoded the messages, the entropy depended on the file itself. For instance a.txt would have far less entropy than aaa.txt since it had only one value character a, whereas aaa.txt had thousands of a's making it so that there was far more data contributing to more uncertainty. This can be shown by the bar graph shown below:




As we can see, aaa.txt has far more entropy than the other files since it contains far more characters. We can also see that after the files were encoded, they had far more entropy than before.

Upon closer look, I decided to see what would happen if errors were introduced and graph how entropy would change.




As we can see the entropy increases as the error rate approaches 50% but then decreases as the error rate increases towards 100%. This is because when errors are initially introduced, they can cause a huge increase in corrected and uncorrectable errors. However, as the error rate goes above 50%, the chances of there being more uncorrectable errors increases which in turn means that there will be more of the same null characters which decreases the uncertainty in the outcome of the program. Therefore, maximum entropy can only be achieved when there is the most diversity and randomness of data.

This is because the more random the data is, the more new information there is. If all of the characters are the same, then there isn't much new information at all.