Kevin Thomas Young PIS: 730009388

Collaborators, rdarden1, sumit97

COMP 590 A1

1. Use your decoder to decode the provided compressed file, producing ASCII English text.

File in data folder named huffdecodefinal.txt

1. Calculate the probability of each symbol in the source ASCII English text file by dividing the

number of occurrences for each symbol by the total symbol count.

For loop divides the number of occurrences (bitcnt) by each symbol (num\_symbols) Console prints out the 256 probabilities,

Most probabilities equal 0.0.

I’ve posted the symbol indexs 40-90 with the most dramatic probabilities:

Prob Answer 40 is 8.695773158583076E-6

Prob Answer 41 is 8.695773158583076E-6

Prob Answer 42 is 0.0

Prob Answer 43 is 0.0

Prob Answer 44 is 0.013285402231683223

Prob Answer 45 is 0.001987853744052091

Prob Answer 46 is 0.010794932799065031

Prob Answer 47 is 1.7391546317166151E-6

Prob Answer 48 is 1.4261067980076244E-4

Prob Answer 49 is 1.0956674179814675E-4

Prob Answer 50 is 6.087041211008153E-5

Prob Answer 51 is 2.6087319475749227E-5

Prob Answer 52 is 3.8261401897765534E-5

Prob Answer 53 is 2.7826474107465842E-5

Prob Answer 54 is 2.4348164844032614E-5

Prob Answer 55 is 3.130478337089907E-5

Prob Answer 56 is 6.434872137351476E-5

Prob Answer 57 is 2.2609010212316E-5

Prob Answer 58 is 1.043492779029969E-4

Prob Answer 59 is 3.513092356067563E-4

Prob Answer 60 is 0.0

Prob Answer 61 is 0.0

Prob Answer 62 is 0.0

Prob Answer 63 is 0.0012817569635751455

Prob Answer 64 is 0.0

Prob Answer 65 is 0.0013234966747363442

Prob Answer 66 is 8.330550685922587E-4

Prob Answer 67 is 5.687035645713332E-4

Prob Answer 68 is 3.7217909118735567E-4

Prob Answer 69 is 4.0870133845340455E-4

Prob Answer 70 is 3.304393800261569E-4

Prob Answer 71 is 2.678298132843587E-4

Prob Answer 72 is 0.0021635083618554694

Prob Answer 73 is 0.006563569580098506

Prob Answer 74 is 1.9826362801569412E-4

Prob Answer 75 is 1.3565406127389598E-4

Prob Answer 76 is 5.06093997829535E-4

Prob Answer 77 is 0.0012730611904165623

Prob Answer 78 is 5.28703008041851E-4

Prob Answer 79 is 5.426162450955839E-4

Prob Answer 80 is 3.182652976041406E-4

Prob Answer 81 is 3.47830926343323E-5

Prob Answer 82 is 3.547875448701895E-4

Prob Answer 83 is 0.0013252358293680607

Prob Answer 84 is 0.001980897125525225

Prob Answer 85 is 8.00011130589643E-5

Prob Answer 86 is 1.478281436959123E-4

Prob Answer 87 is 0.0013182792108411942

Prob Answer 88 is 1.3913237053732921E-5

Prob Answer 89 is 7.913153574310599E-4

Prob Answer 90 is 3.4783092634332303E-6

1. Calculate the theoretical entropy of the source message in bits per symbol using the symbol

probabilities from part 2. What is this value?

Theo Entropy Answer 4.53

1. What is the compressed entropy achieved by the provided compressed file in bits per symbol?

(Note: Do not include the overhead incurred by the 260 header bytes.)

Compressed Entropy Answer 14.16

1. Use your encoder to re-encode the raw English text.

File in bin/data folder named huffdecodefinalcompressed.txt

1. What is the compressed entropy achieved by the Huffman code produced by your encoder?

Compressed Entropy Answer 4.57

1. Using your entropy value calculations as evidence, does your encoder achieve better

compression or worse compression then the original compressed file?

Compressed Entropy Answer 14.16

Compressed Entropy Answer 4.57

Encoder achieves better compression than the decoder.