HW1: Huffman Decoding & Encoding

Ami Zou (PID 730047435)

**Link to my Github Repo:** <https://github.com/ami-zou/COMP590-A1>

**Part 3**

1. Use your decoder to decode the provided compressed file, producing ASCII English text.
   1. Check out <https://github.com/ami-zou/COMP590-A1/blob/master/data/uncompressed.txt> for the uncompressed file.
2. 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.
   1. See next page for details
3. Calculate the theoretical entropy of the source message in bits per symbol using the symbol probabilities from part 2. What is this value?
   1. H = Sum( - P(i) \* log(P(i)) )
   2. Notice: when P(i) = 0, its entropy is 0
   3. The optimal entropy is **4.5317029489157985**
4. 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.)
   1. Entropy = Sum[ code\_length \* count] / total\_number
   2. The given encoder has entropy **5.858789339677769**
5. Use your encoder to re-encode the raw English text.
   1. Check out <https://github.com/ami-zou/COMP590-A1/blob/master/data/new_recompressed.txt> for the new\_recompressed file
6. What is the compressed entropy achieved by the Huffman code produced by your encoder?
   1. Entropy = Sum[ code\_length \* count] / total\_number
   2. My Huffman Encoder has entropy **4.5729401452541945**
7. Using your entropy value calculations as evidence, does your encoder achieve better compression or worse compression than the original compressed file?
   1. As calculated, my encoder has a lower entropy than the given encoder, and therefore it has a better compression
   2. Also, can use `$: ls -l` to show the file details:
      1. Original: 349484
      2. New: 328936
   3. Therefore, my encoder achieves better compression (20,548 shorter).

Total numbers of letters: 574992

2. Probability for each symbol:

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol has probability 0.0

Symbol

has probability 0.021995088627320032

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol

has probability 0.021995088627320032

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol \_ has probability 0.0

Symbol has probability 0.16466837799482428

Symbol ! has probability 6.000083479422323E-4

Symbol " has probability 0.008857514539332721

Symbol # has probability 0.0

Symbol $ has probability 0.0

Symbol % has probability 0.0

Symbol & has probability 8.695773158583076E-6

Symbol ' has probability 0.002598297019784623

Symbol ( has probability 8.695773158583076E-6

Symbol ) has probability 8.695773158583076E-6

Symbol \* has probability 0.0

Symbol + has probability 0.0

Symbol , has probability 0.013285402231683223

Symbol - has probability 0.001987853744052091

Symbol . has probability 0.010794932799065031

Symbol / has probability 1.7391546317166151E-6

Symbol 0 has probability 1.4261067980076244E-4

Symbol 1 has probability 1.0956674179814675E-4

Symbol 2 has probability 6.087041211008153E-5

Symbol 3 has probability 2.6087319475749227E-5

Symbol 4 has probability 3.8261401897765534E-5

Symbol 5 has probability 2.7826474107465842E-5

Symbol 6 has probability 2.4348164844032614E-5

Symbol 7 has probability 3.130478337089907E-5

Symbol 8 has probability 6.434872137351476E-5

Symbol 9 has probability 2.2609010212316E-5

Symbol : has probability 1.043492779029969E-4

Symbol ; has probability 3.513092356067563E-4

Symbol < has probability 0.0

Symbol = has probability 0.0

Symbol > has probability 0.0

Symbol ? has probability 0.0012817569635751455

Symbol @ has probability 0.0

Symbol A has probability 0.0013234966747363442

Symbol B has probability 8.330550685922587E-4

Symbol C has probability 5.687035645713332E-4

Symbol D has probability 3.7217909118735567E-4

Symbol E has probability 4.0870133845340455E-4

Symbol F has probability 3.304393800261569E-4

Symbol G has probability 2.678298132843587E-4

Symbol H has probability 0.0021635083618554694

Symbol I has probability 0.006563569580098506

Symbol J has probability 1.9826362801569412E-4

Symbol K has probability 1.3565406127389598E-4

Symbol L has probability 5.06093997829535E-4

Symbol M has probability 0.0012730611904165623

Symbol N has probability 5.28703008041851E-4

Symbol O has probability 5.426162450955839E-4

Symbol P has probability 3.182652976041406E-4

Symbol Q has probability 3.47830926343323E-5

Symbol R has probability 3.547875448701895E-4

Symbol S has probability 0.0013252358293680607

Symbol T has probability 0.001980897125525225

Symbol U has probability 8.00011130589643E-5

Symbol V has probability 1.478281436959123E-4

Symbol W has probability 0.0013182792108411942

Symbol X has probability 1.3913237053732921E-5

Symbol Y has probability 7.913153574310599E-4

Symbol Z has probability 3.4783092634332303E-6

Symbol [ has probability 0.0

Symbol \ has probability 0.0

Symbol ] has probability 0.0

Symbol ^ has probability 0.0

Symbol \_ has probability 0.0

Symbol ` has probability 0.0

Symbol a has probability 0.059781701310626933

Symbol b has probability 0.010231446698388847

Symbol c has probability 0.01769068091382141

Symbol d has probability 0.03191174833736817

Symbol e has probability 0.09193693129643543

Symbol f has probability 0.015278473439630465

Symbol g has probability 0.013448882767064586

Symbol h has probability 0.04835371622561705

Symbol i has probability 0.045854551019840276

Symbol j has probability 5.878342655202159E-4

Symbol k has probability 0.006026170798898072

Symbol l has probability 0.029311712162951833

Symbol m has probability 0.019226354453627182

Symbol n has probability 0.04935373013885411

Symbol o has probability 0.05773993377299162

Symbol p has probability 0.011568856610178924

Symbol q has probability 7.060967804769458E-4

Symbol r has probability 0.04226841416924062

Symbol s has probability 0.04596933522553357

Symbol t has probability 0.06581308957342015

Symbol u has probability 0.02273075103653616

Symbol v has probability 0.007594888276706458

Symbol w has probability 0.01827503687007819

Symbol x has probability 9.408826557586889E-4

Symbol y has probability 0.015610651974288337

Symbol z has probability 2.556557308623424E-4

Symbol { has probability 0.0

Symbol | has probability 0.0

Symbol } has probability 0.0

Symbol ~ has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol  has probability 0.0

Symbol   has probability 1.7391546317166151E-6

Symbol ¡ has probability 0.0

Symbol ¢ has probability 1.7391546317166151E-6

Symbol £ has probability 0.0

Symbol ¤ has probability 0.0

Symbol ¥ has probability 0.0

Symbol ¦ has probability 0.0

Symbol § has probability 0.0

Symbol ¨ has probability 1.7391546317166151E-6

Symbol © has probability 2.0869855580599383E-5

Symbol ª has probability 0.0

Symbol « has probability 0.0

Symbol ¬ has probability 0.0

Symbol ­ has probability 0.0

Symbol ® has probability 0.0

Symbol ¯ has probability 0.0

Symbol ° has probability 0.0

Symbol ± has probability 0.0

Symbol ² has probability 0.0

Symbol ³ has probability 0.0

Symbol ´ has probability 0.0

Symbol µ has probability 0.0

Symbol ¶ has probability 0.0

Symbol · has probability 0.0

Symbol ¸ has probability 0.0

Symbol ¹ has probability 0.0

Symbol º has probability 0.0

Symbol » has probability 0.0

Symbol ¼ has probability 0.0

Symbol ½ has probability 0.0

Symbol ¾ has probability 0.0

Symbol ¿ has probability 0.0

Symbol À has probability 0.0

Symbol Á has probability 0.0

Symbol Â has probability 0.0

Symbol Ã has probability 2.6087319475749227E-5

Symbol Ä has probability 0.0

Symbol Å has probability 0.0

Symbol Æ has probability 0.0

Symbol Ç has probability 0.0

Symbol È has probability 0.0

Symbol É has probability 0.0

Symbol Ê has probability 0.0

Symbol Ë has probability 0.0

Symbol Ì has probability 0.0

Symbol Í has probability 0.0

Symbol Î has probability 0.0

Symbol Ï has probability 0.0

Symbol Ð has probability 0.0

Symbol Ñ has probability 0.0

Symbol Ò has probability 0.0

Symbol Ó has probability 0.0

Symbol Ô has probability 0.0

Symbol Õ has probability 0.0

Symbol Ö has probability 0.0

Symbol × has probability 0.0

Symbol Ø has probability 0.0

Symbol Ù has probability 0.0

Symbol Ú has probability 0.0

Symbol Û has probability 0.0

Symbol Ü has probability 0.0

Symbol Ý has probability 0.0

Symbol Þ has probability 0.0

Symbol ß has probability 0.0

Symbol à has probability 0.0

Symbol á has probability 0.0

Symbol â has probability 0.0

Symbol ã has probability 0.0

Symbol ä has probability 0.0

Symbol å has probability 0.0

Symbol æ has probability 0.0

Symbol ç has probability 0.0

Symbol è has probability 0.0

Symbol é has probability 0.0

Symbol ê has probability 0.0

Symbol ë has probability 0.0

Symbol ì has probability 0.0

Symbol í has probability 0.0

Symbol î has probability 0.0

Symbol ï has probability 0.0

Symbol ð has probability 0.0

Symbol ñ has probability 0.0

Symbol ò has probability 0.0

Symbol ó has probability 0.0

Symbol ô has probability 0.0

Symbol õ has probability 0.0

Symbol ö has probability 0.0

Symbol ÷ has probability 0.0

Symbol ø has probability 0.0

Symbol ù has probability 0.0

Symbol ú has probability 0.0

Symbol û has probability 0.0

Symbol ü has probability 0.0

Symbol ý has probability 0.0

Symbol þ has probability 0.0

Symbol ÿ has probability 0.0

The root node is 0.9999999999999999 with height 27