CS-150 Worksheet 3 Solution Release date: 22/10/2019

This worksheet is about getting familiar with compression methods and applying them to different pieces of data. **Remember that whitespace characters count in the calculation of the message length!**

□ Task 3 Solution.1

1. **Solution:** Common keywords may vary, but we could choose (for example) the mapping:

Keyword	Encoding
uffalo	@
and	*
ison	+
sentence	=
to	<
in	%

Believe it or not, this = is grammatically correct * has mean%g: “B@ b@ B@ b@ b@ b@ B@ b@.” First devised by professor William J. Rapaport % 1972, the = uses various mean%gs * parts of speech for the term “b@” (* its related proper noun “B@”) < make an extremely hard-<-parse =.

Although most people know “b@” as both a s%gular * plural term for b+, * “B@” as a city % New York, “b@” is also a verb mean%g “< bully, confuse, deceive, or %timidate.” Us%g these def%itions, Wikipedia suggests the = can be read: [Those] (B@ b@) [whom] (B@ b@ b@) b@ (B@ b@).

Still <o hard < follow for those of us who don’t know “b@” as a verb. Ref%e once more: [Those] b@(es) from B@ [that are intimidated by] b@(es) from B@ %timidate b@(es) from B@.

And once more: B+ from B@, New York who are %timidated by other b+ % their community also happen < %timidate other b+ % their community.

2. Calculate the compression ratio of the new compressed message.

Solution: Original message length: 1085, Encoded length: 896.

Ratio: 896/1085 = 0.826 (to 3 decimal places)

□ Task 3 Solution.2

1. Apply Run-Length encoding to the following:

- AAAAAAAAAAaaaAAAABBCCCCDDDdAAAAaEEEEEE

  Solution: *A10aaa*A4BB*C4DDDd*A4a*E6

  Original message length: 38, Encoded length: 26, Ratio = 26/38 = 0.684

- 1011101101110000000101011111101000001111001000001

Solution: 101110110111 *07 1010 *16 01 *05 *14 001 *05 1, obviously here there is then the issue of how you represent the components of the encoding. The flag character identifies the start of the encoding, and the single character that follows is the one to repeat, but the potential for a multi-digit number of repetitions may be ambiguous. In this example I've used spaces to reduce ambiguity, but consider the follow showing the same information by using square brackets: 101110110111*0[7]1010*1[6]01*0[5]*1[4]001*0[5]1. When counting the message length, don't include the spaces or brackets used to reduce ambiguity (they aren't actually part of the message here).

Original message length: 49, Encoded length: 37, Ratio = 37/49 = 0.755

□ Task 3 Solution.3

1. Construct a Huffman Tree and encode the following message:

   - the cat in the hat sat on the mat

2. Calculate the compression ratio of the new compressed messages above.

   Solution: See Huffman_solution.pdf