QR48/CSCI E-2 Assignment 1 Solutions

1.

(a) (3 points) Suppose I send 25 random facts about myself to 25 other people and ask them to wait a day and then do the same thing, sending 25 random facts about themselves to 25 other people. If everybody manages to send their lists to 25 people who haven't been contacted before, how long will this go on before everyone on earth has been contacted? (You can assume that everyone in the world can be reached.)

Answer: According to https://www.cia.gov/library/publications/the-world-factbook/print/xx.html, world population is 6,706,993,152. Hence 25^x=6,706,993,152. Taking log base 25 of both sides (lg(6,706,993,152)/lg(25)), we see that it should take 7.02929153 days, or 8 days total rounded up.

(It did not matter where you got your number for world population as long as you cited where you got it from.)

(b) (3 points) If you are on Facebook, you may have seen some of your friends (perhaps even you!) fill out '25 Random Things About Me', which is then sent on to other friends. According to an article in Time magazine (http://www.time.com/time/arts/article/0,8599,1877187,00.html), approximately 5 million of Facebook's 130 million users sent 25 facts to their friends in just one week.

Assuming that it takes one day for Facebook users who receive a notice from friends to fill out their own random facts, how many Facebook users on average post a new '25 Random Things' out of the 25 users who are invited to do so. (Hint: use the formula for exponential growth.)

Answer: 5 million=x^7, so x=7th root of 5 million= 9, so 9 users on average fill out the 25 random things at each stage.

(c) (3 points) Using your answer from part B (If you can't figure out part B, just pick a number between 1 and 25), answer the following: How long would you expect it to take for all 130 million active Facebook users to fill out a '25 Random Things About Me' note? Explain why your answer might not be reached in practice.

Answer: 130 million =9^x, so x=log base 9 of 130 million= 8.5, so 9 days. Might not happen in practice because eventually 9 people aren't always going to respond, some people will never respond, as time goes on people who have already responded will be among the 25 people, and perhaps there are disconnected groups of users in the Facebook friend network.

2. In this problem, half of the exercise is to find some estimates of the size of things. Whatever estimates you use, state them; the problem is to draw reasonable conclusions from reasonable assumptions, not to get any one "right" answer. Use Google or whatever your favorite search engine is, and explain where you got your numbers.

   (a) (2 points) The new version of the Amazon Kindle can store 1500 books in its 2GB of memory (http://www.pcmag.com/article2/0,2817,2340623,00.asp). At this rate, what size is an average book?

   Answer: 2GB/1500=1.33MB per book

   (b) (3 points) Jeff Bezos, Amazon's founder, has stated that his goal for the Kindle is to make every book available for download. Let's go one step further – how much storage would be required to hold Widener Library?

   Answer: 3 million books according to Wikipedia (http://en.wikipedia.org/wiki/Widener_Library), but could also count the 15.6 million books in the Harvard Library system. Again, any reasonable number was accepted with citation.

   3 million*1.33MB=4TB.

   (c) (3 points) According to "Kryder's Law", disk storage capacity doubles annually. Given that the Kindle is 2GB today, how many years will it take before you can store all of Widener Library on a Kindle-like device if Kryder's Law continues to hold?

   Answer: 4TB=4000GB, so 2*2^x=4000, 2^x+1=4000, so x=lg(4000)-1=10.96 years.

   Many people forgot to account for the growth in the equation starting with a disk of size 2GB, so the calculation was off by a year because without including that it is as if you begin with a disk of size 1GB.

   (d) (2 points) Hard disk drives used in desktop PCs can store far more than mobile devices such as Kindles and iPods. Currently, a one terabyte disk drive can be purchased for around $100. In how many years might you expect to store all of Widener Library on a single hard disk drive?

   Answer: 1*2^x=4, so x=lg(4)= 2 years

3. The members of the Itty Bitty Kitty Committee are stranded on a life raft after their ship sank. They need to send a message out so that they can be rescued, but all they have is a cell phone that has just enough battery power left to transmit a few bytes of data. Afraid that their message will be too long to send, they decide to compress it with a Huffman code. Help the kitties send their message, which you can find in this picture

(a) (3 points) Calculate the entropy of the message. (Hint: use the character frequency and the entropy equation.)

Answer: The message is:

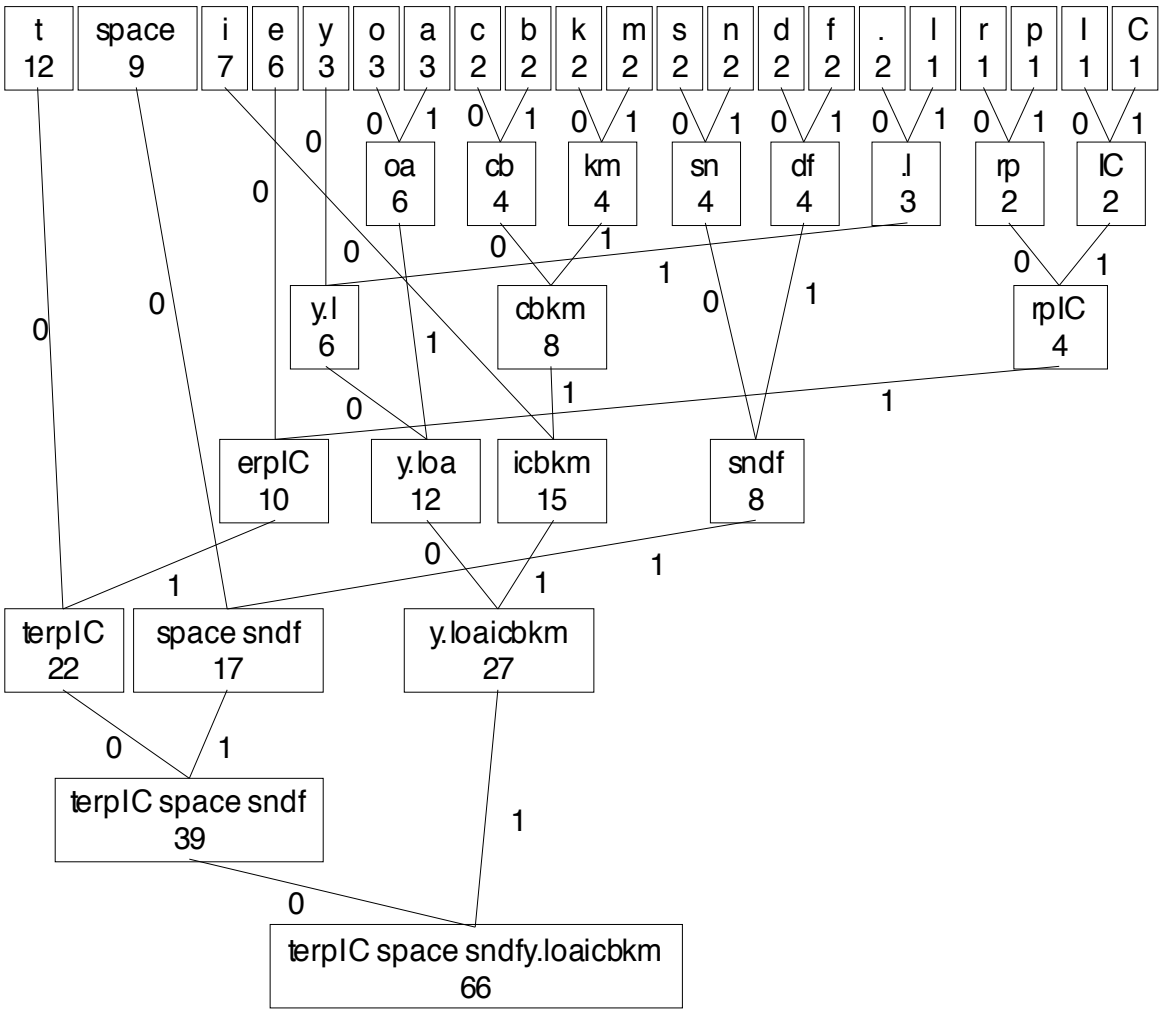
Itty bitty kitty comitte boat sinked. Comitte escaped in liferaft.

Some people did not account for spaces or periods, which is incorrect. They are part of the message. Some ignored the capital/lowercase distinction, which is not ideal but less problematic than deleting characters completely.

The following chart shows the counts, frequencies and self-information. Summing all the self information, we get the entropy of the whole message: 3.98.

Letter	Count	Freq (p)	- p lg p
t	12	0.18	0.45
space	9	0.14	0.39
i	7	0.11	0.35
e	6	0.09	0.32
y	3	0.05	0.2
o	3	0.05	0.2
a	3	0.05	0.2
c	2	0.03	0.15
b	2	0.03	0.15
k	2	0.03	0.15
m	2	0.03	0.15
s	2	0.03	0.15
n	2	0.03	0.15
f	2	0.03	0.15
.	2	0.03	0.15
d	2	0.03	0.15
p	1	0.02	0.09
l	1	0.02	0.09
r	1	0.02	0.09
I	1	0.02	0.09
C	1	0.02	0.09
	66		3.98

(b) (4 points) Create a Huffman code for the message; don't forget to show all of your work.


t 12
space 9
i 7
e 6
y 3
o 3
a 3
c 2
b 2
k 2
m 2
s 2
n 2
d 2
f 2
. 2
l 1
r 1
p 1
l 1
C 1
oa 6
cb 4
km 4
sn 4
df 4
.l 3
rp 2
lC 2
y.l 6
cbkm 8
rplC 4
erplC 10
y.loa 12
icbkm 15
sndf 8
terplC 22
space sndf 17
y.loaicbkm 27
terplC space sndf 39
terplC space sndfy.loaicbkm 66

Letter	**huff code**
t	000
space	010
i	110
e	0010
y	1000
o	1010
a	1011
c	11100
b	11101
k	11110
m	11111
s	01100
n	01101
f	01111
.	10010
d	01110
p	001101
l	10011
r	001100
I	001110
C	001111

(c) (3 points) Encode the message using the Huffman code you created and write down its bit representation. How long is it?

Answer:

Itty : 001110 000 000 1000 010
bitty : 11101 110 000 000 1000 010
kitty : 11110 110 000 000 1000 010
comitte : 11100 1010 11111 110 000 000 0010 010
boat : 11101 1010 1011 000 010
sinked. : 01100 110 01101 11110 0010 01110 10010 010
Comitte : 001111 1010 11111 110 000 000 0010 010
escaped : 0010 01100 11100 1011 001101 0010 01110 010
in : 110 01101 010
liferaft.: 10011 110 01111 0010 001100 1011 01111 000 100010

0011100000001000010111011100000001000010111101100000001000010111001010
11111110000000001001011101101010110000100110011001101111100010011101001
001000111110101111111000000000100100010011001110010110011010010
011100101100110101010011110011110010001100101101111000100010

It is 266 bits long.

(d) (3 points) Using your entropy calculation from part A, what is the efficiency of your compression?

Answer:

average code length: 266/66 = 4.03

efficiency = 3.98/4.03 = 98.7%

Many people failed to use a weighted average when computing average code length for this computation.

(e) (2 points) Convert the message to its binary ASCII representation (remember, you can use tools from the web – there's no need to do this by hand). How long is the message in bits? Is it safe to say that the kitties can communicate using fewer bits using their custom Huffman-encoded transmission when compared to standard ASCII?

Answer:

01001001011101000111010001111001001000000110001001101001011101000111010001111001001000000110001111001001000000110101101101001011101000111010001111001001000000110001101101111011011010110100101110100011101000111100100100000011000100101111011000010111010000100000011100110110100101101110011010110110010001100100001011100010000001000011010111101101101011010010111010001101100100001011100010000001000011010111101101101011010010111010001110010100100000011001010111001101100011011000010111000001100101011001000000110010101110011011000110110000101110000011001010110010000010000001101001011011100010000001101100011010010110011001100101011100100110000101100110011101000010111000100000

It is 537 bits long. It is safe to say that the kittys can commuicate using their custom code (as long as they have some way of communicating the encoding along with the code)

4. A single 43-minute episode of the television show Lost is available for download on iTunes in both standard definition and high definition. The high-definition version is 1.43 gigabytes, and the standard-definition version is 632 megabytes. The same episode sent across the wire on Verizon FiOS TV took 8.17 gigabytes for 62 minutes, including commercials.

(a) (1 point). What size would the FiOS transmission have been had there been no commercials?

Answer: 43/62*8.17GB= 5.67GB

(b) (1 point). Compare the size of the iTunes HD version to the FiOS HD version. How much smaller is the iTunes version?

Answer: the iTunes version is 1.43GB, so 1.43GB/5.67GB=25% of the size of the FiOS version.

(c) (2 points) Verizon FiOS high-speed internet advertises a download speed of up to 20 megabits per second. Assuming you were able to get that maximum download rate continuously, how long would it take you to get the high-definition version of the episode from iTunes? What about the standard definition version?

Answer: Assume 1GB=1024MB (they could also assume that 1GB=1000MB, but they need to state their assumption).

20 megabits per second= 20 megabits / 8bits per byte=2.5 megabytes per second.

2.5 megabytes per second times x = 1.43GB=1.43*1024 megabytes=1464 megabytes, so x =1464/2.5=585 seconds, or 9 minutes 45 seconds for the HD version.

Standard def=632 MB/2.5 MB=252 seconds, or 4 minutes, 12 seconds.

(d) (3 points) The truth is that real-world speeds rarely approach the maximums that are advertised. A more typical speed is 7 megabits per second. At a continuous rate of 7 megabits per second, how long would it take to get the high-definition version of the episode from iTunes? What about the standard definition version? And finally, what about the television version?

Answer: 7 megabits per second=0.875 megabytes per second

iTunes HD: 1464 MB/0.875MB/s=1673 s

iTunes SD: 632 MB/0.875MB/s=722 s

FiOS HD: 5.67GB*1024MB/GB=5806 MB, so 5806MB/0.875MB/s= 6635 s

5. (Graduate students: 6 points, undergraduate students: 0 points) In 1950, Shannon estimated the entropy of the English language by the following experiment. He reduced the language to 27 characters, the 26 letters of the alphabet plus space, by converting upper case to lower case and any punctuation mark to space. Starting from a random position in a long text, he showed a subject 100 consecutive characters and asked the subject to state what the 101st is. If the subject didn't get it right the first time, ask again; this can't go on more than 27 times. Repeat the experiment 100 times. If it takes r tries pr percent of the time, an estimate of the maximum information content of the text is $\sum_r p_r \log(1/p_r)$ bits per character.

   Try a modified version of this experiment. Go to http://www.eecs.harvard.edu/qr48/applets/ent.html. There you will have a choice of

several texts. For each trial, it will show about 30 characters, and you need to pick the next one. Do 20 trials each on two separate texts and write a few (no more than 5) sentences about your experience and your results. (The web page figures the results for you; the formula above is for your information only.) As a point of reference, Huffman coding gets you to about 4.7 bits per character. (Note that some of the teaching staff experienced problems with the applet---letters not appears, B and M keys switched, etc.) If this applet does not work for you, please use the alternative applet available at http://www.math.ucsd.edu/~crypto/java/ENTROPY/. For this applet you will be asked to guess an entire quote starting from the first letter. Do this 5 times and then write a few (no more than 5) sentences about your experience and rsults. Be sure to indicate which applet you used in your answer.

Answer: All answers that demonstrated use of the applets and offered observations on the experience received credit for this problem.