QR 48 Midterm Examination (*50 points total*)

1. (*19 points total*) Suppose you have to encode five symbols A, B, C, D and E. Each symbol occurs this frequently:
   A: 5
   B: 1
   C: 1
   D: 20
   E: 20

   (a) (2 points) How many bits would it take to encode the symbols using a fixed-length coding scheme?

   (Answer: there are 5 outcomes, so 3 bits)

   (b) (4 points) What is the smallest possible average number of bits per symbol for any code for this source?

   (Answer: -20/47*lg(20/47)+-20/47*lg(20/47)-1/47*lg(1/47)-1/47*lg(1/47)-5/47*lg(5/47) = 1.6293 bits

   (c) (5 points) Construct a Huffman code for the symbols. Show your Huffman tree and the resulting code.

   (d) (2 points) Explain in one sentence the importance of your Huffman code being a prefix code.

   (Answer: Because no code word is a prefix or initial substring of any other code word, prefix codes can be read, unambiguously, left to right.)

   (e) (3 points) Suppose we must encode the string 'DEDE'. How many bits are required using the fixed-length scheme described in part (a)? How many bits are required using your Huffman code derived in part (c)?

   (Answer: Using fixed length encoding, 3*4=12 bits are required. Using Huffman code, 1+2+1+2=6 bits required.)

   (f) (3 points) How efficient is the fixed-length coding scheme? How efficient is your Huffman code?

   (Answer: 1.6293/3=54% efficiency for fixed-length

   3*5/47+4*1/47+4*1/47+2*20/47+1*20/47=1.766 bits on average for Huffman, so 1.6293/1.766=92.26% efficiency)

2. (*5 points total*) You can type English text using just the eight buttons of a telephone keypad that have letters on them, even though each key represents three or four letters. The software is smart enough, after you have typed ahead a bit, to figure out which letters your previous keystrokes were supposed to represent. What does this fact alone tell you about the entropy of English?

   (Answer: The entropy of English can't exceed lg 8, or 3 bits per character.)

3. (*12 points total*) My first digital camera used a 720 KB floppy disk to store images, and that disk held only 30 photos. The resolution of cameras (and hence the size of each image file) has doubled every 20 months since then.

   (a) (3 points) Suppose the increase in resolution had not been accompanied by any change in storage technology. When would the size of a single image file have exceeded the capacity of the floppy disk?

   (Answer: Doubling every 20 months, to be 30X larger you need about lg(30)=4.906 cycles. 5*20 months=100 months from 1994, or mid-2002. You can calculate it close enough to be considered correct by successive doubling 24KB until it reaches 720KB – a bit less than 5 cycles, then multiplying that by 20 months.)

(b) (4 points) Suppose instead that the memory capacity of a camera has doubled every 16 months since 1994. How much memory will a camera have in 2010?

(Answer: Image size in 2010 should be 720KB * 2^(16*3/4) = about 2.8125 GB. You can get there by doubling as well, so an understanding of logs and exponents isn't entirely necessary: 192 months between 1994 and 2010, so doubles 192/16=12 times. 720KB*2^12=2.8125GB.

You may also interpret the difference as 17 years (start of 1994 to end of 2010). In this case

720KB * 2^(17*3/4)=4.73GB or 4959808KB)

(c) (5 points) In reality, an image size of 5MB is as good as film, so there is no reason in ordinary photography to make the image files any larger. For video at 30 frames per second, we can get by with resolution only a quarter of that per frame. How many minutes of video could I shoot with a 2010 camera?

(Answer: 1.25MB/frame*30frames/second*x seconds=2.8125GB*1024MB/GB, so x=76.8 seconds.)

4. (*9 points total*)

(a) (3 points) In one sentence, describe what Jonathan Zittrain means by the Internet's *generativity*?

(Answer: an open platform enables the creation of new applications developed by people other than the system's designers.)

(b) (6 points) Describe, with an example, *two* threats to the Internet's generative future (one or two sentences each) identified by Zittrain.

(Answer: I had in mind these two examples. (1) The move to information appliances, such as the iPhone, threatens innovation by limiting which applications can be run on devices; (2) information insecurity, exemplified by malware that is used to infect PCs for sending spam and committing identity theft, may drive people to reject generative platforms in preference to controlled information appliances. Other examples might include the undermining of net neutrality, because by favoring one application over another, you necessarily limit the generative capability of the unfavored app. Another example is censorship, such as when Pakistan unilaterally took down YouTube by fidgeting with Internet routing tables, which can undermine confidence in keeping platforms open to generativity.)

5. (*5 points total*) Google's main source of revenue comes from paid advertisements that appear alongside normal search results.

(a) (2 points) What company pioneered this business model?

(Answer: Overture, can also mention the website goto.com)

(b) (3 points) How did the company's approach differ from Google's?

(Answer: Overture sold ads for all the search results, not just off to the side separate from the organic results.)