CSE 490 G
Assignment 4
Due Friday, February 3, 2006

1. Let us try LZW on a special class of inputs too. Again assume the two symbol alphabet $\{a, b\}$. Consider the following strategy for encoding the dictionary symbols from LZW. Start with a dictionary of size 2 and use just one bit to transmit a symbol. When the dictionary fills up we double its size to 4 and use two bits to transmit a word in the dictionary. This doubling happens when ever the dictionary fills.

   (a) Encode a^6 and a^{28} with this version of LZW.

   (b) Compute the length, as a function of n, of the encoding of a^n with this version of LZW. (You may restrict yourself to easy n's to work with if that helps.)

   (c) Encode a^6 and a^{28} using the γ-code to represent the dictionary symbols from LZW on the strings a^6 and a^{28}.

   (d) Compute the length, as a function of n, of the encoding of a^n using the γ-code to represent the dictionary symbols of LZW. (You may restrict yourself to easy n's to work with if that helps.)

2. In this example we explore Sequitur on some pathological strings.

   (a) Use Sequitur to find the grammar for a^4, a^8, a^{16}.

   (b) Generalize to give a grammar for a^n for n a power of 2.

   (c) Assuming a two letter alphabet compute the compression ratio for Sequitur, as a function of n, for strings of the form a^n where n is a power of 2. Use the encoding describe on slide 44 of the lectures.

3. In this example we explore LZ77 on pathological strings. For this problem we assume LZ77 (solution B), that is, the search buffer has size the length of the string and the look-ahead buffer has size 0.

   (a) Use LZ77 (solution B) to find the sequence of tuples for a^4, a^8, a^{16}.

   (b) Generalize to give the sequence of tuples for a^n for n a power of 2.

   (c) Assuming a two letter alphabet compute the compression ratio for LZ77 (solution B), as a function of n, for strings of the form a^n where n is a power of 2. Use the simple fixed length code for this.