

CS698D II Sem 2013-14
Homework 2

1. Implement LZ78 compression algorithm using a trie or a suffix-tree data structure. **[30 points]**
2. Plot the graph of the compressibility of the first n bits of the Ehrenfeucht-Mycielski sequence, as n grows from 1 to 10,000 using the program that you have written. The first 10,000 bits of the Ehrenfeucht-Mycielski sequence can be obtained at

<http://barnyard.syracuse.edu/mseq/mseq.shtml>

You can use gplot to plot the graph.

[20 points]

3. Construct a non-normal binary sequence in which every string appears as a phrase in the LZ78 parsing. Show that the sequence constructed is not normal. **[10 points]**
4. Let the diluted Champernowne sequence be defined by

$$c = 00\ 01\ 000\ 001\ 010\ 011\ 0000\ \dots,$$

that is, by preceding every phrase in the Champernowne sequence with a 0. Assuming that the Champernowne sequence has LZ78 compressibility 1, compute the LZ78 compressibility of the diluted Champernowne sequence. **[15 points]**