

# CSEP 521 - Spring 2005

## Assignment 7

Due 5/19/05

1. Consider the string “eta\_ceta\_and\_beta\_ceta” where the blank counts as a symbol.
  - (a) Find the context-free grammar for this string produced by Sequitur.
  - (b) Compute the length of the simple code for this string assuming an initial alphabet of  $\{a, b, c, d, e, n, t, _\}$ . See slide 73.
2. Consider the string “eta\_ceta\_and\_beta\_ceta” where the blank counts as a symbol which is last in the symbol ordering. The string is indexed 0 to 21.
  - (a) Do a most significant symbol first radix sort (bucket sort) to order the cyclic shifts of the string. Initially, there are 8 buckets one for each of  $\{a, b, c, d, e, n, t, _\}$ . Each number 0 to 21 ends up in a bucket according to the first symbol in the cyclic shift starting at that index. A bucket is subdivided further if it has more than one element in it. This demonstrates that only linear space is needed to sort the cyclic shifts.
  - (b) From the result in (a) compute the  $L$  and  $X$  in the Burrows-Wheeler transform.
  - (c) Use move-to-front coding of  $L$  to create a symbol stream which can be entropy coded.
  - (d) Compute the first-order entropy of the resulting symbol stream.