CSEP 590
Assignment 5
Due November 8, 2007

1. 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 based on the second symbol in the cyclic shift if it has more than one element in it. This continues until all buckets have at most one element in them. 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.

2. Decode the following using the Burrows-Wheeler transform algorithm. $L = baaaaaa$ and $X = 3$. In the process compute the mapping T and use it in the decoding. Show the steps along the way.