582487 Data Compression Techniques (Spring 2012)
Exercises 4 (15 February)

Solve the following problems before the exercise session and be prepared to present your solutions at the session.

1. Let $T_k = X^{n/k}$, where X is a random (uncompressible) string of length k over a binary alphabet. What is the (asymptotic) compression ratio of the following algorithms on T_k?

   (a) LZ77 with $d_{\max} = l_{\max}$ and fixed length encoding
   (b) LZ77 without length or distance limits and γ coding

2. Let T_k be as in Problem 1. What is the (asymptotic) compression ratio of LZ78 on T_k?

3. Encode the following string using LZFG:

   ```
   how much wood would a woodchuck chuck
        if a woodchuck could chuck wood
   ```

4. Let R be the string of terminals and non-terminals resulting from running Re-Pair on a text T. Let α and β be two substrings of R. Show that

   $$\text{expand}(\alpha) = \text{expand}(\beta) \text{ if and only if } \alpha = \beta$$

   where $\text{expand}(\alpha)$ is the result of repeatedly replacing non-terminals in α with their right-hand side until there are no non-terminals left.

5. Let $\Sigma = \{\$, \_, a, b, c, \dots, z\}$ with the ordering $\$ < \_ < a < b < \cdots < z$.

   (a) Decode `lnmthm_caaaain$iiin`.
   (b) Encode `lnmthm_caaaain$iiin` using MTF encoding. The initial list contains the symbols in alphabetical order.

6. (a) Compute the BWT of `abcabcabc` without the sentinel symbol $.
   (b) Compute the inverse BWT assuming the position of the last text symbol has been provided. What problem do you encounter? How can you solve it?