582487 Data Compression Techniques (Spring 2012)
Exercises 3 (8 February)

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

1. Let $T = \texttt{baadadaada}$ and $\Sigma = \{\texttt{a}, \texttt{b}, \texttt{c}, \texttt{d}\}$. Encode T using adaptive Huffman coding.

2. Let $T = \texttt{senselessness}$. Compute $H_k(T)$ for all k such that $H_k(T)$ is positive.

3. Let $n_1, n_2, \ldots, n_\ell$ be a sequence of non-negative integers summing up to n. Interpolative coding encodes the sequence as follows:

   - Split the sequence into two halves: $n_1, \ldots, n_{\lfloor \ell/2 \rfloor}$ and $n_{\lfloor \ell/2 \rfloor + 1}, \ldots, n_\ell$.
   - Encode the sum of the first half using the shortest fixed length code that can encode all possible values.
   - Encode each half recursively.

   Use interpolative coding to encode all symbol frequencies needed for semiadaptive first order compression on the text $\texttt{senselessness}$ over the alphabet $\{\texttt{e}, \texttt{l}, \texttt{n}, \texttt{s}\}$.

4. Show that

$$nH_0(T) = n \log n - \sum_{s \in \Sigma} n_s \log n_s$$

$$nH_k(T) = \sum_{w \in \Sigma^k} n_w \log n_w - \sum_{w \in \Sigma^{k+1}} n_w \log n_w$$

   Does the above mean that $\sum_{k=0}^{\infty} H_k(T) = \log n$ for all T?

5. (a) What is the length of the longest string with exactly z phrases in its LZW parsing?

   (b) What is the length of the shortest *binary* string with exactly z phrases in its LZW parsing?

6. Decode $\langle 0, \texttt{o} \rangle$ $\langle 0, \texttt{t} \rangle$ $\langle 0, \texttt{a} \rangle$ $\langle 0, \texttt{␣} \rangle$ $\langle 0, \texttt{v} \rangle$ $\langle 3, \texttt{l} \rangle$ $\langle 2, \texttt{a} \rangle$ $\langle 4, \texttt{v} \rangle$ $\langle 6, \texttt{t} \rangle$ $\langle 3, \texttt{v} \rangle$ $\langle 9, \texttt{a} \rangle$ $\langle 8, \texttt{a} \rangle$ $\langle 0, \texttt{l} \rangle$ $\langle 2, \texttt{i} \rangle$ $\langle 1, \texttt{l} \rangle$ $\langle 2, \texttt{a} \rangle$.