- This problemset has *three* questions.
- For other questions, either send the solutions to gawry1+aos@gmail.com, or leave them in the envelope attached to the doors of my office (room 321).

1. Huffman code is defined as follows: given the probability $p(c)$ of each character $c \in \Sigma$ we want to select all $\text{code}(c) \in \{0,1\}^+$ so that they are prefix-free, meaning that no $\text{code}(c)$ is a proper prefix of $\text{code}(c')$, and that the expected cost $\sum_{c\in\Sigma} p(c)|\text{code}(c)|$ is minimized.
   (a) Construct the Huffman code for $\Sigma = \{a, b, c, d, e\}$ and the probabilities $p(a) = \frac{12}{31}$, $p(b) = \frac{6}{31}$, $p(c) = \frac{5}{31}$, $p(d) = \frac{4}{31}$, $p(e) = \frac{4}{31}$. Compute the corresponding expected cost.
   (b) Alphabetical Huffman code have the additional property that $\text{code}(c) < \text{code}(c')$ if $c < c'$, where the first inequality is understood as the lexicographical comparison. Show an example where this additional restriction increases the expected cost.

2. Recall that the zeroth order entropy was defined as $H(w[1..n]) = \sum_{c\in\Sigma} \frac{n_c}{n} \log \frac{n}{n_c}$, where n_c is the number of occurrences of the letter c in $w[1..n]$.
   (a) Compute the entropy for $w = a^{12}b^6c^5d^4e^4$.
   (b) Prove that if the number $\ell_1, \ell_2, \ldots, \ell_k$ are chosen so that $\sum_i 2^{-\ell_i} \leq 1$, it is possible to construct a binary tree on k leaves, where each ℓ_i is a depth of one of the leaves.
   Hint: This is known as (one half) of the Kraft's inequality. Use induction on k.
   (c) Show that the Huffman cost is less than $H(w[1..n]) + 1$. You might find the previous subquestion useful.
   Hint: show that some code with the expected cost not exceeding $H(w[1..n]) + 1$ exists, then argue that the Huffman code is the best possible, so can't be worse.

(SPOJ) 3. Solve the BWHEELER problem, which asks you to reverse the Burrows-Wheeler transform.