58093 String Processing Algorithms (Autumn 2010)
Exercises 4 (2 December)

1. Let $\mathcal{R} = \{\texttt{manne}, \texttt{manu}, \texttt{minna}, \texttt{salla}, \texttt{saul}, \texttt{sauli}, \texttt{vihtori}\}$.
   (a) Give the compact trie of $\mathcal{R}$.
   (b) Give a balanced ternary tree of $\mathcal{R}$.
2. Let $\mathcal{R}$ be as in Problem 1. Give a balanced binary tree of $\mathcal{R}$ with precomputed lcp information.
3. What is the time complexity of the Aho–Corasick algorithm when σ is not constant using
   (a) array implementation
   (b) binary tree implementation
   (c) hash table implementation.

   Choose the implementation details to minimize the time complexity.
4. Modify ternary search tree to support prefix queries in time $\mathcal{O}(|S| + \log n + |\mathcal{Q}|)$, where S is the query string and $\mathcal{Q}$ is the result of the query.
5. Show that the worst case time complexity of string binary search without precomputed lcp information is $\Omega(m \log n)$.
6. Define

$$LCP[mid] = \max\{LLCP[mid], RLCP[mid]\}$$
$$L[mid] = \begin{cases} 1 & \text{if } LCP[mid] = LLCP[mid] \\ 0 & \text{otherwise} \end{cases}$$

   Show that, if we store the arrays LCP and L instead of $LLCP$ and $RLCP$, we can compute $LLCP[mid]$ and $RLCP[mid]$ when needed during the string binary search.