1. [3+3+3 points] Each of the following pairs of concepts are somehow connected. Describe the main connecting factors or commonalities as well as the main separating factors or differences.

   (a) Morris–Pratt algorithm and Crochemore algorithm.
   (b) Shift–And algorithm and Myers’ bitparallel algorithm.
   (c) LCA (Lowest Common Ancestor) preprocessing and RMQ (Range Minimum Query) preprocessing.

   A few lines for each part is sufficient.

2. [10 points] Construct the Aho–Corasick automaton for the pattern set \{angel, angry, chapel, gel, michael\}. Simulate the scanning of the text michelangelo with the automaton.

3. [4+7 points] A $q$-gram of a string is its factor of length $q$. Let $G_q(A, B)$ denote the number of $q$-grams shared by the strings $A$ and $B$.

   For example, for $A = \text{varaurat}$ the 2-grams are va, ar, ra, au, ur, ra and at, and for $B = \text{ararat}$ they are ar, ra, ar, ra and at. The shared 2-grams are ra twice, ar and at, and thus $G_2(A, B) = 4$.

   (a) Show that if $ed(A, B) \leq k$, then $G_q(A, B) \geq |A| - q + 1 - kq$.
   (b) Design a filtering algorithm for approximate string matching based on the result of (a)-part.

4. [4+6 points]

   (a) What is the lcp-comparison technique? Describe the main principles.
   (b) Give two examples of algorithms that use the lcp-comparison technique. Describe the role of the lcp-comparison technique in the algorithms.

5. [10 points] Let $S$ and $T$ be strings over an alphabet of constant size. Describe an algorithm that given $S$, $T$ and an integer $k$ finds if there exists a string that occurs exactly $k$ times in $S$ and exactly $k$ times in $T$. The time complexity should be linear.