Practice Exercises for Final

1. **Listing the $k$ smallest numbers in sorted order**
   
   Given an unsorted list $S$ of $n$ numbers and an integer $k \leq n$, design an efficient algorithm to list the $k$ smallest numbers in $S$ in sorted order. Analyze the worst-case running time of your algorithm.

2. **LCS of three sequences**
   
   Give an efficient algorithm to determine the longest common subsequence of three sequences $X$, $Y$, and $Z$ of length $m$, $n$, and $p$, respectively. Analyze the worst-case running time of your algorithm.

3. **Basic graph structures**
   
   For each of the following statements, indicate whether it is true or false. Briefly justify your answers.

   (a) If $T$ is a minimum spanning tree of a weighted undirected graph $G$, then the unique path connecting any two vertices $u$ and $v$ in $T$ is a shortest path between $u$ and $v$ in $G$.

   (b) If $T$ is the depth-first search tree rooted at a node $r$ of an unweighted undirected graph $G$, then the path connecting $r$ to any vertex $v$ in $T$ is a shortest path between $r$ and $v$ in $G$.

4. **Alternating paths**
   
   You are given a directed graph $G = (V, E)$ in which each vertex has been assigned a color, either red or blue. A directed path in $G$ is called an **alternating red-blue path** if and only if no two consecutive vertices on the path have the same color. Give an efficient algorithm that determines for all pairs of vertices $u, v$ in $V$ whether $v$ is reachable from $u$ via an alternating red-blue path. Briefly justify the correctness of your algorithm and analyze its worst-case running time.

5. **Hamiltonian path**
   
   A **Hamiltonian path** of a directed graph $G$ is a simple path in $G$ that visits every vertex in $G$ exactly once. Design a linear time algorithm to determine whether a given **directed acyclic graph** has a Hamiltonian path. (Hint: Use topological sort.)

6. **Data compression**
   
   - You have two data sets $foo$ and $bar$, each having a million characters from the alphabet \{a, b, c, d\}. The probability distribution of characters in $foo$ is even (0.25 for each character), while that in $bar$ is (0.5, 0.25, 0.125, 0.125). Which data set is more compressible?
   
   - Give the LZ77 (basic Lempel-Ziv compression) code for the string $abababababcab$, assuming a dictionary of size 5 and a lookahead buffer of 4.