系所組別：資訊管理研究所乙組

考試科目：資料結構 考試日期：0223，節次：3

※ 考生請注意：本試題不可使用計算機。 請於答案卷(卡)作答，於本試題紙上作答者，不予計分。

1. [10%] Let $f(n)$ and $g(n)$ be asymptotically positive functions. Answer True or False for each of the following conjectures. (2% each)
   a. $f(n) = o(g(n))$ if and only if $g(n) = \omega(f(n))$
   b. $2^n = \Omega(n!)$
   c. Every n-node B-tree has height $O(log(n))$
   d. QuickSort has a worst case running time of $O(nlog(n))$
   e. Queues can be implemented using stacks

2. [10%] Snake is one of the classic video games on Nokia mobile phones around the 1990s. The player controls a dot on a bordered plane. As it moves, it leaves a trail behind, so that it resembles a moving snake. Each time the snake eats an object, it becomes one unit longer. Now, suppose you are a software engineer working at HTC and your boss asked you to implement this game on an Android phone.
   a. Name a data structure that can be used to represent a snake and explain why. (4%)
   b. Explain in detail how to manipulate the data structure to simulate the movement of a snake (6%)

3. [10%] Given the preorder and postorder traversals of a full binary tree, write a recursive procedure to reconstruct the binary tree.

4. [15%] Autocomplete is a feature supported in all major browsers. As the user types a letter, it predicts and displays possible options to fill in a field, based on earlier typed words or phrases.
   a. Name a data structure that can be used to implement such a feature and explain why. (4%)
   b. Suppose the set of earlier typed words are: iPad, iPod, iPad2, MBA, MBP. Show how these strings can be stored in your data structure and explain what options will be displayed when the user enters "iP". (6%)
   c. What is the time complexity, in general, to build such a data structure? Explain why. (5%)

5. [5%] Given a database containing a large number, say n, of patients' records. Suppose you are a doctor and you wanted to find m ($m \ll n$) patients who have higher blood pressure readings than the others to participate in a new drug test. Unfortunately, the computer memory can only hold $m + 1$ records regardless of how they are stored in the memory. Write a pseudo algorithm to solve this problem in $O(nlog(m))$ time. You may assume all the patients in the database have different blood pressure readings.

（背面仍有題目，請繼續作答）

※ 考生請注意：本試題不可使用計算機。 請於答案卷(卡)作答，於本試題紙上作答者，不予計分。

6. [28%] True or False, and EXPLAIN
Circle T or F for each of the following statements to indicate whether the statement is true or false, respectively. If the statement is correct, briefly state why. If the statement is wrong, explain why or give a counter example. Answers without reasons will get at most 1 point.

(a) (**T** , **F**) To construct a binary min heap of n nodes, one requires $\Omega(n \log n)$ time since there are n objects to be inserted and each insertion takes $O(\log n)$ time to update the heap.

(b) (**T** , **F**) Given two n-element arrays A and B, it takes $\Omega(n^2)$ time to check whether there exists a common element u between A and B since one requires two for loops to compare each element of A to each element of B.

(c) (**T** , **F**) For a directed connected simple graph $G = (V, E)$ with no directed cycle, even if some arc has negative length, any shortest path can be calculated in $O(|E|)$ time, which is faster than the Dijkstra's method.

(d) (**T** , **F**) Given a bipartite simple undirected graph that contains more than 400 edges, its minimum spanning tree always contains the edges of the smallest, the 2nd smallest, and the 3rd smallest weights.

(e) (**T** , **F**) A strongly connected directed graph must contain a directed Euler circuit.

(f) (**T** , **F**) Suppose you are asked to store the IDs for those customers renting or returning a bike in a station of C-Bike, the first public bike sharing system in Taiwan. Among heap, linked list, and array, linked list is more suitable for this job.

(g) (**T** , **F**) In a final exam, there are n problems to be solved within time limit T, where problem i requires t_i time to solve and get g_i points. Suppose you can always give a correct answer if you spend t_i time for problem i, but $\sum_{i=1}^{n} t_i > T$, which means you could NOT finish answering all n problems within the time limit T. In order to get the highest grade, you design a polynomial time greedy algorithm that selects the problem i of larger $\frac{g_i}{t_i}$ in sequence. This greedy algorithm guarantees to get the highest grade within $O(n \log n)$ time since it requires sorting $\frac{g_i}{t_i}$ in nonincreasing order.

7. [10%] Please fill out (a1),...(a5); (b1),...,(b5) by **y** (as yes) or **n** as (no)

	counting sort	heap sort	merge sort	quick sort	insertion sort
sorted in place	(a1)	(a2)	(a3)	(a4)	(a5)
stable sorting	(b1)	(b2)	(b3)	(b4)	(b5)

※ 考生請注意：本試題不可使用計算機。 請於答案卷(卡)作答，於本試題紙上作答者，不予計分。

8. [12%] In a social network $G = (V, E)$, where each vertex $i \in V$ represents an individual i and each edge $(i, j) \in E$ between vertices i and j represents their acquainted relation. Using facebook as an example, suppose you are given a set of $|V| = n$ facebook IDs, denoted by $i = 1, ..., n$, and there are totally $|E| = m$ mutual relations between all individuals. An $n \times n$ relation matrix $R[i][j]$ records whether i is a friend of j In particular, $R[i][j] = 1$ indicates i and j are mutual friends, and $R[i][j] = 0$ indicates i and j are strangers. Note that two individuals are either friends or strangers. For example, if ID[3], ID[5] and ID[6] are friends of ID[2], then $R[2][3] = R[2][5] = R[2][6] = 1$ (all other $R[2][j] = 0$ for $j \in \{1, 2, ..., n\} - \{3, 5, 6\}$). Of course the relation is symmetric, i.e. $R[i][j] = R[j][i]$ for each $i, j \in V$. To avoid confusion, we define the diagonal entry $R[i][i] = 0$ for each $i \in V$. For simplicity, assume G is composed by several component which contains no isolated vertex.

Answer the following questions with explanation. (answers without explanation get at most 1 point.)
(a) [6%] Let Cmin[s][t] denote the minimum number of connections between individuals s and t. In other words, Cmin[2][5]=2 means individual 2 and 5 have a common friend; If Cmin[2][5]=∞, it means individual 2 and 5 will never have any connection. For simplicity, let Cmin[i][i]=0 for each individual i.
Give a method that calculates the $n \times n$ matrix Cmin[$\cdot$][$\cdot$]. Explain why your method works and analyze its complexity.

(b) [6%] A so-called "Six degree of separation" is the theory that everyone is connected to each other by at most six connections. Suppose you are asked to test this theory over the given G, which are known to be connected (i.e. only one component). Let Cmax[i] = $\max_{j \in V}$\{Cmin[i][j]: Cmin[i][j]$<\infty$\} denote the maximum number of connections required for individual i to connect to his/her "farthest" individual. Assuming G contains no isolated vertex, which means Cmax[$\cdot$]$<\infty$.
Give a method that takes the $n \times n$ matrix Cmin[$\cdot$][$\cdot$] (i.e. as introduced in (a)) as given input, and uses Cmin[$\cdot$][$\cdot$] to calculate the average of Cmax[$\cdot$] (i.e. $\sum_{i=1}^{n}$ Cmax[i]/n). Explain why your method works and analyze its complexity.