Discrete Math in CS	**Problem Set 5**
CS 280 Fall 2000	**Due October 20, 2000**

Reading: Rosen, Sections 1.4 – 1.6 and 4.1 – 4.3.

As in previous problem sets, some of the questions present "claims," and ask you to decide whether they are true or false. For such a question, you should do one of the following two things: (i) give a proof of the claim; or (ii) give a proof that the claim is false, by showing that its negation is true. (Using a counter-example or other means.)

---

(1) For each of the following claims, do one of the following two things: (i) prove that it holds for all subsets A, B, and C of the natural numbers; or (ii) give an example of subsets A, B, and C of the natural numbers for which it does not hold.

(a) $(A - B) \cup (B - C) = (A \cup B) - C$.

(b) $(B - A) \cap (C - A) = (B \cap C) - A$.

---

(2) We say that a collection of non-empty subsets of $\{1, 2, \ldots, n\}$ is *diverse* if no two of them have any elements in common — in other words, if they are all pairwise disjoint.

Let $d(n)$ denote be the maximum size of a diverse collection of non-empty subsets of $\{1, 2, \ldots, n\}$ — that is, the maximum number of subsets in such a collection. Give a formula for $d(n)$ in terms of n, and prove it is correct. (That is, for every n, show that if your formula specifies $d(n) = d'$, then there exists a diverse collection of subsets of $\{1, 2, \ldots, n\}$ of size d', and there is no diverse collection of size $d' + 1$.)

(3) Consider a computer system with a set S of n independent processors, labeled $1, 2, \ldots, n$. At any given point in time, some subset of these processors will be *available* to perform tasks.

We want to avoid a situation in which at some point in time, a set $A \subseteq S$ is available to perform a certain task; and later, when we return to continue the task, a set $B \subseteq S$ is available, where B is *disjoint* from A. The problem here is that, since no processor in B also belongs to A, there is no processor with memory of what happened previously in processing the task.

We'd like a situation where each pair of "available sets" is guaranteed to have at least one processor in common — such a processor can serve as a "bridge" between consecutive groups that perform a common task, to maintain consistency. In the study of distributed operating systems, such a collection of available sets is sometimes called a *quorum system*.

Thus, formally, we say that a *quorum system of order* n is a collections of subsets of $\{1, 2, \ldots, n\}$ with the property that every pair of sets in the collection has at least one element in common. Let $q(n)$ denote be the maximum size of a quorum system of order n. Give a formula for $q(n)$ in terms of n, and prove it is correct. (That is, for every n, show that if your formula specifies $q(n) = q'$, then there exists a quorum system of order n with q' sets, and there is no quorum system of order n with $q' + 1$ sets.)

(4) Suppose you're analyzing purchase data for a Web site that sells books. They have a catalog consisting of a set B of N books, and they've assigned each book one of g "genres" (e.g. biography, mystery, science fiction, and so forth). Thus, they've partitioned their set B into subsets $G_1, G_2, \ldots, G_g$, where G_i denotes the books in the i^{th} genre. For simplicity, let's suppose that each set G_i has size exactly m, where $N = gm$.

The company is interested in sequences of purchases by a single customer. Thus, we'll say that a *pattern* is a sequence of distinct books from B (representing a sequence of purchases of books – no two of them the same – by a single customer). The *length* of a pattern is simply its length as a sequence.

To decide whether their genre classification system is useful, they are interested in studying what they call "focused patterns." Formally, we define a *focused pattern* to be a pattern of length k in which the final $k-1$ books all belong to the same genre. (I.e. they all come from a single set G_i; the first book in the sequence may or may not come from this genre.)

Let $T(k)$ denote the total number of patterns of length k, and let $F(k)$ denote the total number of focused patterns of length k. A question that arises is: how many of all possible patterns of length k are focused?

Notice that the smallest k for which this definition is at all "interesting" is $k = 3$. Write a formula for $F(3)$ in terms of N, g, and m (not all of these variables need to appear in the formula), and also write a formula for the ratio $F(3)/T(3)$. For each of these two formulas, give a brief explanation for why the formula is correct.

(5) In addition to studying purchase histories, we can study the relationships among purchases made by different customers in database. As discussed in class, a very common way of doing this is to identify each customer i in a database with the *set* of items S_i that he or she has purchased. We will suppose throughout that each set S_i is non-empty.

In terms of this representation, here are two potentially interesting collections of customers. First, we say that a collection D of k customers is *diverse* if there is no item that has been purchased by more than one of the customers in D. Notice that this is the same sense of "diverse" as in Question (2): if the customers in D are represented by sets $S_1, S_2, \ldots, S_k$, then they are diverse if these k sets are pairwise disjoint.

Second, we say that a collection R of k customers is *representative* if, for *every* customer j in the database, at least one of the items purchased by j has also been purchased by a customer in the collection R. Thus, the set associated with each j in the database must have a non-empty intersection with at least one of the sets representing a customer in R.

Intuitively, there is natural juxtaposition between diverse sets and representative sets. Here is a claim that attempts to make this precise.

> *Claim: Consider a customer database in which each customer is represented by a set of items purchased. Let $k \geq 2$ be any natural number. If the database does not contain a diverse set of k customers, then it contains a representative set of ℓ customers, for some natural number $\ell < k$.*

Decide whether you think the claim is true or false, and give a proof of either the claim or its negation.