Homework # 4

Handed out: 9/21/06
Due: 9/26/06

1. Problem 29 from Chapter 2 of Gusfield:
   Suppose we are given a tree where each edge is labeled with one or more characters, and we are given a pattern P. the label of a subpath in the tree is the concatenation of the labels on the edges in the subpath. The problem is to find all subpaths of paths starting at the root that are labeled with pattern P. Note that although the subpath must be part of a path directed from the root, the subpath itself need not start at the root. Give an algorithm for this problem that runs in time proportional to the total number of characters on the edges of the tree plus the length of P.

2. Problem 12 from Chapter 3 of Gusfield:
   Prove that the search phase of the Aho-Corasick algorithm runs in O(m) time if no pattern in P is a proper substring of another, and otherwise in O(m + k) time, where k is the total number of occurrences.

3. Give an example showing that k, the number of occurrences in T of patterns in set P, can grow faster than O(n + m). Be sure you account for the input size n. Try to make the growth as large as possible.