Assignment 12

Post Date: 21 Jan 2009  Due Date: 28 Jan 2009, 14:30
You are permitted and encouraged to work in groups of two.

Problem 1: Naive String Matching

(a) Suppose that all characters in the pattern are different. Show how to accelerate the naive string-matching algorithm such that it runs in $O(n)$ time on an $n$-character text.

(b) Find a text and a pattern such that the naive string-matching algorithm needs $O((n - m + 1) \cdot m)$ time even if the pattern does not occur in the text.

(c) Suppose that pattern $P$ and text $T$ are randomly chosen strings of length $m$ and $n$, respectively, from the $d$-ary alphabet $\Sigma = \{0, 1, \ldots, d-1\}$, where $d \geq 2$. Show that the expected number of character-to-character comparisons in every step made by the naive string-matching algorithm is

$$\frac{1 - d^{-m}}{1 - d^{-1}} \leq 2.$$

Problem 2: Spurious Hits

How many spurious hits does the Rabin and Karp algorithm encounter when looking at the pattern $P = 26$ in $T = 3141592653589793$ with $q = 11$?

Problem 3: Extension of Rabin and Karp

Show how to extend the algorithm of Rabin and Karp to look for an $m \times m$ pattern among $n \times n$ characters.

Problem 4: Wildcards

Now, a pattern can contain also wildcards $\ast$, i.e. a wildcard can stand for arbitrarily many characters. Consider the pattern $P = aba \ast bab$ and the input alphabet $\Sigma = \{a, b, c\}$.

Find a string-matching automaton such that the algorithm for string matching with finite automata finds $P$ in a text $T$. 