Algorithms and Data Structures

Drishti Maharjan

May 18, 2019

Assignment 12

Problem 12.1

Find code in horse.py

Problem 12.2

a. Rabin Karp algorithm matches the hash value of the pattern with the hash value of current substring of text, and if the hash values match then only it starts matching individual characters. We begin by calculating hash values of the pattern of length M, and all subsequent sub strings of length M in text of length N. The hash function we choose is the most important step here, as we must choose it wisely so as to not lead to collisions. The hash function suggested by Rabin and Karp calculates an integer value. The integer value for a string is numeric value of a string.

Let's say the number of possible characters would be d = 256. Let q = 11, a prime number so that chances of collision are less with modulo arithmetic.

Let our hashing function be h = (h * d) mod q, where h initially assigned as 1. Let's take an example where text= ABDCB, and pattern = DC. Then, we have M = 2, N = 5. We get h = 3.

For a window of text of length M, we apply
for (i = 0; i < M; i++)
p = (d * p + pattern[i]) mod q
t = (d * t + text[i]) mod q
end for

While calculating the hash value of the window, in first iteration, we get p = 2, and t = 10, and in second, p = 7, and t = 8. Since, M = 2, p = 7 is the final hash value of our pattern DC, and the hash value of current window text AB = 8. t and p don't match, so we move to next window, by removing trail character and adding leading character. This is done by
t = (d*(t - text[i]*h) + text[i+M]) mod q.

If we have a negative t, we make it positive by
t = t + q

In our next window BD, we get t = -9 + 11 = 2, which is not equal to p = 7. So, we move to next window DC, t = -4 + 11 = 7. Now, t equals p = 7, so we check character by character for the window and pattern. We get DC = DC, so pattern is found at position 3.

Then, checking window by window continues for the rest of the text with CB, and the program ends as it reaches end of text. Since, patterns were not found in any other positions, the program execution ends.

b. Find code in rabin.c