Homework 2 of Sequence Informatics 2008

Dae-Ki Kang

October 2, 2008

Deadline: October 18, 2008

1. Give a linear time $(O(|T|))$ algorithm to determine if a single pattern P occurs as a subsequence in a text T.

2. Give an efficient algorithm to determine if any pattern in a set occurs as a subsequence in a text T. If n is a length of all patterns in the set and m is a length of T, then $O(nm)$ time is obtained by solving the problem for each pattern separately. Try for a time bound that is significantly better than $O(nm)$.

3. A traceback path is the backtracking path that follows the pointers to reconstruct the alignment in the dynamic programming matrix. Since the traceback paths in a dynamic programming table correspond one-to-one with the optimal alignments, the number of distinct optimal alignments can be obtained by computing the number of distinct traceback paths. Give an algorithm to compute the number of optimal alignments in $O(nm)$ time. ($n = |S_1|, m = |S_2|$) (Hint: Use dynamic programming.)

4. Given two strings S_1 and S_2 and a text T, you want to find whether there is an occurrence of S_1 and S_2 interwoven (without spaces) in T. For example, the strings *abac* and *bbc* occur interwoven in *cabbabccdw*. Give an efficient algorithm for this problem. (It may have a relationship to the longest common subsequence problem.)

5. A palindrome is a string that reads the same forwards and backwards. 'A parindrome in a string S' is both a substring of S and a parindrome. A maximal parindrome s in a string S is not shorter than any other parindromes in S. Give a linear-time algorithm (1) to find all maximal parindromes in a string S, (2) to find all maximal even-long parindromes in a string S, (3) and to find all maximal odd-long parindromes in a string S.

6. Show how to count the number of distinct substrings of a string T in $O(m)$, where $m = |T|$.

7. Construct a Huffman tree for the following set of frequencies, based on the first 8 Fibonacci numbers.

```
a:1, b:1, c:2, d:3, e:5, f:8, g:13, h:21
```

Can you generalize your answer to find an optimal code when the frequencies are the first n Fibonacci numbers?