Algorithmic Bioinformatics I: Exercises
Assignment 10

Deadline: Tuesday, 14.07.2009, 10 ct

Exercise 1 (Double Digest):
Develop an algorithm in pseudocode which finds all valid solutions of the double-digest-problem (see lecture slides) for given multisets $A$, $B$ and $C$. Do not forget to state the pseudocode for used auxiliary functions. Comment your pseudocode in detail! Analyse the runtime of your algorithm.

Exercise 2 (PDP):
The partial-digest-problem has been presented in the lecture. Given the multiset

$$A = \{2, 3, 4, 5, 6, 6, 7, 9, 11, 11, 13, 14, 16, 17, 20\}$$

of $\binom{6}{2}$ pairwise distances, reconstruct the positions of the restriction sites with the described procedure. State all intermediate steps as well as the final solution.

Solution: 0, 4, 6, 11, 17, 20

Exercise 3 (PDP, programming):
Implement the corrected version of the algorithm for solving the PDP problem as described in the lecture. Your program has to accept a filename as input parameter (see our homepage for an example). The solution (positions of the restriction sites) should be printed to StdOut.

Exercise 4 (Pattern Matching):
Given are two strings $P$ and $T$ with lengths $m$ and $n$. We are searching for a position $i$ where $P$ occurs in $T$, i.e. where $\forall j, 1 \leq j \leq m : T_{i-1+j} = P_j$.

(a) The “naive” algorithm for pattern matching compares for each position $i$ the pattern $P$ with $T_i \ldots T_{i+m-1}$ until a mismatch occurs or the end of the pattern is reached. If the end is reached, position $i$ is marked as an occurrence of $P$ in $T$. Formulate this algorithm in pseudocode and analyze the runtime.

(b) Assume no character occurs more than once in $P$. How can the naive algorithm be changed such that only $O(n)$ character comparisons are necessary.