Sequence Analysis SS 2014

Freie Universität Berlin, Institut für Informatik
Knut Reinert, David Weese
Sommersemester 2014

7. Exercise sheet, July 3rd, 2014
Discussion: July 8th and 10th, 2014

Exercise 1.

Prove the following q-gram lemma:

Let P and S be strings of length w with at most k mismatches. Then P and S share at least $w + 1 - (k + 1)q$ common q-grams.

Exercise 2.

SWIFT algorithm:

- Sketch the Function $U(n, q, \epsilon)$ for increasing values of n. Let $q = 7$, $\epsilon = 0.1$ and $n_0 = 30$. Draw it for the intervall $n = n_0 \dots n_0 + 35$.
- The Lemma 2.2 in the SWIFT script contains a formula to compute w (for a $w \times e$ parallelogram) Show that every local alignment with τ q-hits and e errors lies in a $w \times e$ parallelogram.

Exercise 3.

Suffix filters:

Determine **all** strong matches for the given weights and edit distances:

$$
\begin{aligned}
i &: 0\ 1\ 2\ 3\ 4\ 5 \\
t_i &: 2\ 1\ 1\ 2\ 1\ 1 \\
dist(A_i, B_i) &: 1\ 1\ 0\ 2\ 0\ 1
\end{aligned}
$$

Exercise 4.

Factor filters:

Prove Theorem 3 (optimal factorization) from the script.