RNA-sequencing: Exercise 2

Sebastian Thieme, Hannes Hauswedell

May 27, 2010

1 PEX

1.1 Lemma

Think about the Lemma shown in the lecture and show that Occ includes a substring that matches p^i with $\lfloor \frac{a_ik}{A} \rfloor$ errors.

1.2 Hierarchical approach

Given a pattern: annual and a text: any_annealing. Build the hierarchical tree and search for possible matching positions indexed by the last index in the text. Do it step by step!

2 q-gram counting

2.1 q-gram lemma

Show that two sequences s_1 and s_2 with length w share at least

$$t = w - q - kq + 1$$

common q-grams (sub-sequences of length q) if their edit distance is $\leq k$. Look at the hints from the lecture if you are stuck!

2.2 gapped q-grams

In the lecture we have shown that the q-gram-lemma is not tight for gapped q-grams. Find a gapped q-gram $(q \ge 4)$ where the actual threshold is higher than that computed by the q-gram-lemma and t > 0. You can basically do this by try and error (repeat the steps from the lecture to find the real threshold).

Good luck!