Programming Homework 2 Instructions

In a practical, we saw Python code implementing the Boyer-Moore algorithm. Some of the code is for preprocessing the pattern P into the tables needed to execute the bad character and good suffix rules — we did not discuss that code. But we did discuss the code that performs the algorithm given those tables:

def boyer\_moore(p, p\_bm, t):

    """ Do Boyer-Moore matching. p=pattern, t=text,

        p\_bm=BoyerMoore object for p """

    i = 0

    occurrences = []

    while i < len(t) - len(p) + 1:

        shift = 1

        mismatched = False

        for j in range(len(p)-1, -1, -1):

            if p[j] != t[i+j]:

                skip\_bc = p\_bm.bad\_character\_rule(j, t[i+j])

                skip\_gs = p\_bm.good\_suffix\_rule(j)

                shift = max(shift, skip\_bc, skip\_gs)

                mismatched = True

                break

        if not mismatched:

            occurrences.append(i)

            skip\_gs = p\_bm.match\_skip()

            shift = max(shift, skip\_gs)

        i += shift

    return occurrences

**Measuring Boyer-Moore's benefit.** First, download the Python module for Boyer-Moore preprocessing:

<http://d28rh4a8wq0iu5.cloudfront.net/ads1/code/bm_preproc.py>

This module provides the BoyerMoore class, which encapsulates the preprocessing info used by the boyer\_moore function above. Second, download the provided excerpt of human chromosome 1:

<http://d28rh4a8wq0iu5.cloudfront.net/ads1/data/chr1.GRCh38.excerpt.fasta>

Third, implement versions of the naive exact matching and Boyer-Moore algorithms that additionally count and return (a) the number of character comparisons performed and (b) the number of alignments tried. Roughly speaking, these measure how much work the two different algorithms are doing.

For a few examples to help you test if your enhanced versions of the naive exact matching and Boyer-Moore algorithms are working properly, see these notebooks:

* [Naive](http://nbviewer.ipython.org/github/BenLangmead/ads1-hw-examples/blob/master/hw2_naive_with_counts.ipynb) : https://nbviewer.org/github/BenLangmead/ads1-hw-examples/blob/master/hw2\_naive\_with\_counts.ipynb
* [Boyer-Moore](http://nbviewer.ipython.org/github/BenLangmead/ads1-hw-examples/blob/master/hw2_bm_with_counts.ipynb): https://nbviewer.org/github/BenLangmead/ads1-hw-examples/blob/master/hw2\_bm\_with\_counts.ipynb