

≡ Item Navigation

Programming Homework 2 Instructions (Read First)

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:

```
1 def boyer_moore(p, p_bm, t):
2     """ Do Boyer-Moore matching. p=pattern, t=text,
3         p_bm=BoyerMoore object for p """
4     i = 0
5     occurrences = []
6     while i < len(t) - len(p) + 1:
7         shift = 1
8         mismatched = False
9         for j in range(len(p)-1, -1, -1):
10             if p[j] != t[i+j]:
11                 skip_bc = p_bm.bad_character_rule(j, t[i+j])
12                 skip_gs = p_bm.good_suffix_rule(j)
13                 shift = max(shift, skip_bc, skip_gs)
14                 mismatched = True
15                 break
16             if not mismatched:
17                 occurrences.append(i)
18                 skip_gs = p_bm.match_skip()
19                 shift = max(shift, skip_gs)
20             i += shift
21     return occurrences
22
```

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](#)
- [Boyer-Moore](#)