CSCI 4192 Assignment 1

Instructor: Travis Gagie

1. With sequencing by synthesis, are errors more likely at the beginning of the read, in the middle, or at the end, or equally likely throughout? Why?

2. What are the pros and cons of short and long reads?

3. Write pseudo-code for Boyer-Moore and give an example (not the same as Ben's) of how it works.

4. How is the Pigeonhole Principle used in approximate matching with an index for exact matching?

5. For the final cost matrix Ben presented for approximate matching, write a sentence explaining each distinct cost.