Name: Parth Motibhai Patel
Nau id: pp594/ 6293553

Writeup Hw2:

Usage for the main file:
./executable <genome_file> <queries_file> <sorted/unsorted> <integer fragment_count/ ALL>

argument 1: genome_file → file path to the genome file
argument 2: queries_file → file path to the queries file
argument 3: sorted/unsorted→ Whether to sort the queries file or not, will also decide what search algorithm to use linear search or binary search.
argument 4: fragment_count/ALL → how many fragments to search from the genome file into the query file.

Eg:
./executable /common/contrib/classroom/inf503/genomes/human.txt /common/contrib/classroom/inf503/human_reads_2.fa unsorted 10000

Runtimes:
    Sorted:
        10K: 0.04706 secs.
        100K: 0.44075 secs.
        1M: 3.40402 secs.
        ALL(3B): 15250.86331

    Unsorted:
        10K: 3796.90595 secs
        100K: 37295.36873 secs
        1M: 5 days. (estimate)
        ALL(3B): 140 days. (estimate)

Sample output for sorted:

[pp594@wind ~/testhome]$ ./executable /common/contrib/classroom/inf503/genomes/human.txt /common/contrib/classroom/inf503/human_reads_2.fa sorted 10000 | tee "debug_output.txt"
Genome Length: 3057186663
Reading queries...
Total genome queries: 31930886
fragment_count: 10000
sorted

For query number: 13707510

Match found:CTAACCCTAACCCTAACCCTAACCCTAACCCT at index: 0 in the genome sequence.

For query number: 22862194
Match found:TAACCCTAACCCTAACCCTAACCCTAACCCTA at index: 1 in the genome sequence.

For query number: 1448914
Match found:AACCCTAACCCTAACCCTAACCCTAACCCTAA at index: 2 in the genome sequence.

For query number: 3912487
Match found:ACCCTAACCCTAACCCTAACCCTAACCCTAAC at index: 3 in the genome sequence.

For query number: 12537253
Match found:CCCTAACCCTAACCCTAACCCTAACCCTAACC at index: 4 in the genome sequence.

For query number: 12762297
Match found:CCTAACCCTAACCCTAACCCTAACCCTAACCC at index: 5 in the genome sequence.

For query number: 13707510
Match found:CTAACCCTAACCCTAACCCTAACCCTAACCCT at index: 6 in the genome sequence.

For query number: 22862194
Match found:TAACCCTAACCCTAACCCTAACCCTAACCCTA at index: 7 in the genome sequence.

For query number: 1448914
Match found:AACCCTAACCCTAACCCTAACCCTAACCCTAA at index: 8 in the genome sequence.

For query number: 3912487
Match found:ACCCTAACCCTAACCCTAACCCTAACCCTAAC at index: 9 in the genome sequence.
Runtime: 0.04706 secs.

Sample output for unsorted:

[pp594@wind ~/testhome]$ ./executable
/common/contrib/classroom/inf503/genomes/human.txt
/common/contrib/classroom/inf503/human_reads_2.fa unsorted 10000 | tee "debug_output.txt"
Genome Length: 3057186663
Reading queries...
Total genome queries: 31930886
fragment_count: 10000
sorted

For query number: 1341754
Match found:CTAACCCTAACCCTAACCCTAACCCTAACCCT at index: 0 in the genome sequence.

For query number: 1083374
Match found:TAACCCTAACCCTAACCCTAACCCTAACCCTA at index: 1 in the genome sequence.

For query number: 15015919
Match found:AACCCTAACCCTAACCCTAACCCTAACCCTAA at index: 2 in the genome sequence.

For query number: 1438593
Match found:ACCCTAACCCTAACCCTAACCCTAACCCTAAC at index: 3 in the genome sequence.

For query number: 4628971
Match found:CCCTAACCCTAACCCTAACCCTAACCCTAACC at index: 4 in the genome sequence.

For query number: 4131680
Match found:CCTAACCCTAACCCTAACCCTAACCCTAACCC at index: 5 in the genome sequence.

For query number: 1341754
Match found:CTAACCCTAACCCTAACCCTAACCCTAACCCT at index: 6 in the genome sequence.

For query number: 1083374
Match found:TAACCCTAACCCTAACCCTAACCCTAACCCTA at index: 7 in the genome sequence.

For query number: 15015919

Match found:AACCCTAACCCTAACCCTAACCCTAACCCTAA at index: 8 in the genome sequence.

For query number: 1438593
Match found:ACCCTAACCCTAACCCTAACCCTAACCCTAAC at index: 9 in the genome sequence.
Runtime: 3796.90595 secs.