

Aim: In this assignment, given two sequences T and P we ask to find whether P occurs exactly within T , and if it does, the locations of P in T . Your program should implement the following algorithms:

- Brute force search
- Knuth-Morris-Pratt
- Boyer-Moore
- Rabin-Karp

Input: Two strings T and P , where $|T| \geq |P|$. These two strings will be given in two files in FASTA format¹. Note that FASTA file allows a single string to be represented in multiple lines.

Example (text.fa):

```
>text
TAAGTCTATACCATCGTAGTCTAATTAACGTTATGGTAGGAT
ATCAAGGACGGAATGACCGCAGAGGCGACGTTAATGCGCCGT
CAGAGACGCCCTAAAGATTGCGGTAGGGTCCCGTTGTTAAAG
AGACTTGAGTGGGTGCTTGATGGGAGTGTATTAAGGGCATGT
ATAAGTGTGCTGGGTCTAAGGCATTAAAGCTGAGTCAATAG
TTACATTGCAGATTAACGAGATCTGAAATTAAGGGAGAGATT
CCCAGAGTGGCCTAGTACTTAAGGGCACCCACGCCGCAGGCG
GCCCTACGCCCGTTAATGGTTCGAGTGCTATTCACTAACACA
TTAACGGACGTTTAGTGTGGATTATAGGTGAAGGGTCTGCGC
CACTCCAAGGCAGGGAACATATGTGTTGTTACTATCTTAACG
```

Example (pattern.fa):

```
>pattern
TGGGTCTAAGGCATTAAAGCTGAGTCAATAGT
```

Output: For each of the four algorithms to be implemented, report:

- Whether P is in T , and if it is, the location of P within T (1-based coordinate).
- Number of character comparisons performed.
- Run time in microseconds.

Finally, report the algorithm that performed the best.

Notes:

- You must write your code yourself. Sufficient evidence of plagiarism will be treated the same as for plagiarism or cheating.
- Non-compiling submissions will not be evaluated.
- Your code must be complete.
- Do not submit the program binary. You must submit the following items:
 - All of the source files
 - A script to compile the source code and produce the binary (Makefile).
 - A README.txt file that describes how the compilation progress works.
- Submit your answers through the Moodle page.
- Use 'CS481 Assignment-1' in the subject line of your e-mail.
- Zip your files and send them in only one zipped file. File name format='surname_name.hw1.zip'
- C / C++, Python, Java will be used as programming language. STL is allowed.
- All submissions must be made by 23:59, October 20, 2019.
- The overall fastest implementation wins. **Bonus** will be given for the fastest code.

¹https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=BlastHelp