

# Assignment Project Exam Help Assignment Tutorial

<https://powcoder.com>

MSBD5009/COMP5112 Parallel Programming  
Assignment 1: Super-mer Generation with MPI

Add WeChat powcoder

# Tutorial Overview

- Problem Description
- Implementation Instruction
- Environment Setup

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

# Problem Description

- Basic Concepts

1. Read

- A DNA fragment with base A, C, T, G (i.e. a string contains 'A', 'C', 'T', 'G' only).

- CAAATTACTGCATA

2. K-mer

- A length- $k$  substring on a read. A read of length  $n$  contains  $n - k + 1$  k-mers.
- ( $k=9$ ) CAAATTACT, AAATTACTG, ..., TACTGCATA are the k-mers of the above read

3. Minimizer

4. Super-mer

.

<i>Read</i> =	CAAATTACTGCATA
(k-mer #1)	<u>CAAATTACT</u>
(k-mer #2)	<u>AAATTACTG</u>
(k-mer #3)	<u>AATTACTGC</u>
(super-mer #1)	<u>CAAATTACTG</u> super-mer #1 is made up of k-mer #1 and #2, minimizer
(k-mer #4)	<u>ATTACTGCA</u>
(super-mer #2)	<u>AATTACTGC</u> super-mer #2 is made up of k-mer #3 only, minimizer
(k-mer #5)	<u>TTACTGCAT</u>
(k-mer #6)	<u>TACTGCATA</u>
(super-mer #3)	<u>ATTACTGCATA</u> super-mer #3 is made up of k-mer #4 #5 #6, minimizer

# Problem Description

- Basic Concepts

1. Read

2. K-mer

3. Minimizer

- The lexicographically smallest length- $p$  substring of a  $k$ -mer.
- ( $p=5$ ) The minimizer of **CAAATTACTG** is **AAATT**.

4. Super-mer

- A substring of a read generated by merging multiple consecutive  $k$ -mers which have the same minimizer value.
- ( $k=9, p=5$ ) The first super-mer in the read **CAAATTACTGCATA** will be **CAAATTACTG** because the first two  $k$ -mers have the same minimizer **AAATT**.

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

<i>Read</i> =	<b>CAAATTACTGCATA</b>
(k-mer #1)	<b><u>CAAATTACT</u></b>
(k-mer #2)	<b><u>AAATTACTG</u></b>
(k-mer #3)	<b><u>AATTACTGC</u></b>
(super-mer #1)	<b><u>CAAATTACTG</u></b> super-mer #1 is made up of k-mer #1 and #2, minimizer is <b>AAATT</b>
(k-mer #4)	<b><u>ATTACTGCA</u></b>
(super-mer #2)	<b><u>AATTACTGC</u></b> super-mer #2 is made up of k-mer #3 only, minimizer is <b>AAATT</b>
(k-mer #5)	<b><u>TTACTGCAT</u></b>
(k-mer #6)	<b><u>TACTGCATA</u></b>
(super-mer #3)	<b><u>ATTACTGCATA</u></b> super-mer #3 is made up of k-mer #4 #5 #6, minimizer is <b>TTACT</b>

# Problem Description

- Your Task

- Input

- Many reads
    - *Given in CSR format in the program*

- Output

- All the super-mers generated from these reads
    - *You need to save all the super-mers to a vector of strings "all\_supermers" in Process 0*

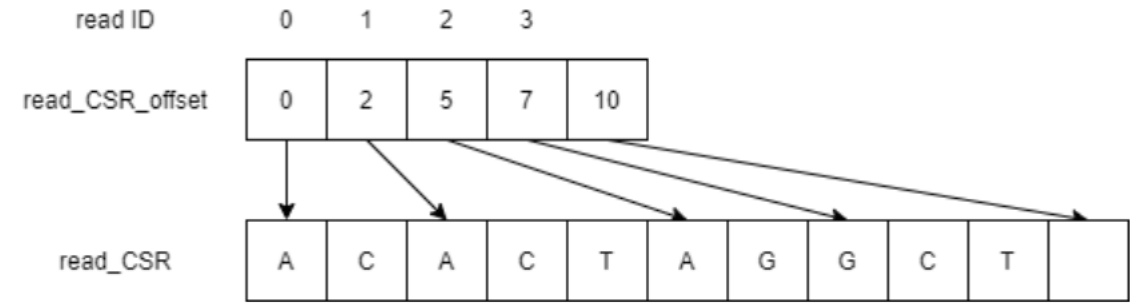


Figure 2: An Example of CSR Format

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

```
// Input data
int num_of_reads = 0;
char* reads_CSR;
/**/int* reads_CSR_offs;
```

```
// Output data, save all the supermers
vector<string> all_supermers;
```

# Implementations

- The code skeleton ***gensuper-mer\_mpi.cpp***
  - Already implemented:
    - MPI initialization and finalization
    - Loading reads from the dataset file and converting to CSR format
    - Result correctness checking
    - Outputting super-mers to text file
    - \* Function ***read2supermers***(...) which can convert a read to its corresponding super-mers
  - You need to:
    - Scatter the read data to each MPI process
    - Perform the super-mer generation in each process
      - You can refer to the sequential version to know the usage of the function `read2supermers(...)`
    - Gather all the super-mers to Process 0 and store in the vector "all\_supermers"
      - Each string represents a super-mer
      - The order in the vector doesn't matter

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

# Implementations

- Only write your code in the specified area of ***gensuper-mer\_mpi.cpp*** and only submit this file to Canvas.

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

```
// =====  
// =====  
// ==== write your implementation only below this line ====  
// =====  
  
// =====  
// ==== write your implementation only above this line ====  
// =====  
// =====
```