

### 3L Construct a String Spelled by a Gapped Genome Path

---

#### Gapped Genome Path String Problem

Reconstruct a string from a sequence of  $(k, d)$ -mers corresponding to a path in a paired de Bruijn graph.

**Input:** A collection of  $(k, d)$ -mers  $(a_1 | b_1), \dots, (a_n | b_n)$  such that  $\text{SUFFIX}(a_i | b_i) = \text{PREFIX}(a_{i+1} | b_{i+1})$  for all  $i$  from 1 to  $n - 1$ , *PairedReads*.

**Output:** A string *Text* where the  $i$ -th  $k$ -mer in *Text* is equal to  $\text{SUFFIX}(a_i | b_i)$  for all  $i$  from 1 to  $n$ , if such a string exists.

```
TAA-GCC
AAT-CCA
ATG-CAT
TGC-ATG
GCC-TGG
CCA-GGG
CAT-GGA
ATG-GAT
TGG-ATG
GGG-TGT
GGA-GTT
TAATGCCATGGGATGTT
```

---

#### Formatting

**Input:** Space-separated integers  $k$  and  $d$  followed by a space-separated list of paired  $k$ -mer strings *PairedReads* where individual  $k$ -mers within the pair are separated by a "|" character.

**Output:** A string *Text* where the  $i$ -th  $k$ -mer in *Text* is equal to  $\text{SUFFIX}(a_i | b_i)$  for all  $i$  from 1 to  $n$ , if such a string exists.

#### Constraints

- The value of  $k$  will be between 1 and  $10^2$ .
- The value of  $d$  will be between 1 and  $10^3$ .
- The number of strings in *PairedReads* will be between 1 and  $10^4$ .
- The length of any one pair of paired  $k$ -mers in *PairedReads* will be between 1 and  $10^2$ .
- All  $k$ -mer strings in *PairedReads* will be DNA strings.

## Test Cases

### Case 1

---

**Description:** The sample dataset is not actually run on your code.

**Input:**

4 2  
GACA|TCTC ACAC|CTCT CACA|TCTC ACAT|CTCA

**Output:**

GACACATCTCTCA

### Case 2

---

**Description:** The sample dataset is not actually run on your code.

**Input:**

2 1  
AC|TT CG|TG GT|GA TT|AC

**Output:**

ACGTTGAC

### Case 3

---

**Description:** The sample dataset is not actually run on your code.

**Input:**

2 4  
GC|CG CA|GT AT|TG TA|GC AC|CA CC|AT

**Output:**

GCATACCGTG CAT

### Case 4

---

**Description:** The sample dataset is not actually run on your code.

**Input:**

5 1  
ACAGC|GCGAA CAGCT|CGAAT AGCTG|GAATC GCTGC|AATCA

**Output:**

ACAGCTGCGAATCA

### Case 5

---

**Description:** A larger dataset of the same size as that provided by the randomized autograder. Check input/output folders for this dataset.