

2.2 Problem 2

Step 1

Reconstruction of Bacteria

Here, you are given **two different problems: the solution for the hard version is not necessarily accepted in the easy version**. The problems are very similar but they have a difference in the provided information: the easy version has an additional information compared to the hard one. For each problem the tests are provided. You should submit only answers in the plain text format. Each test file contains a number of subtests. Since this problem has an exact solution your score for the test is calculated using the percentage of the correct answers. For example, if there are T subtests and your answer is correct for K of them then your score for the test will be $P \cdot K/T$, where P is the total score for the test.

For the first test the maximum score is 600, while for the second — 1100.

Note: The tests data does not represent the real-life data. So, we suggest not to rely solely on biological knowledge.

Step 2

Reconstruction of Bacteria

Our hapless scientist John Doe has big problems at work. Again... The PI gave him a bowl with bacteria and asked him to calculate the number of different bacteria species that live there. For that John brings the bowl into sequencing laboratory. Unfortunately, the equipment is quite old and it only returns a list of different k -mers of bacteria genomes that are in the bowl.

He decided that he can estimate the number of different bacteria species in the bowl as the minimal number of bacteria genomes such that each k -mer appears in at least one genome and the genomes do not contain k -mers other than provided by the sequencing facility. Help him find that number!

Note 1: In the easy version of the problem it is known that each k -mer appears in the genome of exactly one bacteria and only once.

Note 2: In this problem we consider the DNA strand of the bacteria as a linear strand (not cyclic) and the k -mers are taken only in one direction.

Input format

Each test contains a number of subtests separated by a newline.

The first line of a subtest contains two integers n and k ($1 \leq n, k \leq 100$)— the number of different k -mers that were read and their length.

Each of the next n lines of the subtest contains one k -mer — a string of length k consisting of letters A, C, G and T.

Output format

The output should contain the answers for all subtests. The answers should be separated by a newline.

The first line of the answer should contain m — the minimal number of different bacteria species. The next m lines should contain the genome sequences of these bacteria. Each sequence cannot be shorter than k and cannot be longer than 10 000.

If you do not have an answer for the test you could return 0 for it instead.

It is guaranteed that the answer with this constraints always exists.

Note: the examples are provided in the corresponding steps.

Step 3

Reconstruction of Bacteria (Easy version)

Download the test at <https://stepik.org/media/attachments/lesson/33452/test1.zip>.

Sample input

```
5 3
ACT
TCT
CTA
TAT
TAG
```

Sample output

```
3
ACTAG
TCT
TAT
```

To solve this problem please visit <https://stepik.org/lesson/33452/step/3>

Step 4

Reconstruction of Bacteria (Hard version)

Download the test at <https://stepik.org/media/attachments/lesson/33452/test2.zip>.

Sample input

```
5 3
ACT
TCT
CTA
TAT
TAG
```

Sample output

```
2
ACTAG
TCTAT
```

To solve this problem please visit <https://stepik.org/lesson/33452/step/4>