

1969번 - DNA

시간 제한	메모리 제한	제출	정답	맞은 사람	정답 비율
2 초	128 MB	884	478	412	57.865%

문제

DNA란 어떤 유전물질을 구성하는 분자이다. 이 DNA는 서로 다른 4가지의 뉴클레오티드로 이루어져 있다(Adenine, Thymine, Guanine, Cytosine). 우리는 어떤 DNA의 물질을 표현할 때, 이 DNA를 이루는 뉴클레오티드의 첫글자를 따서 표현한다. 만약에 Thymine-Adenine-Adenine-Cytosine-Thymine-Guanine-Cytosine-Cytosine-Guanine-Adenine-Thymine로 이루어진 DNA가 있다고 하면, "TAACTGCCGAT"로 표현할 수 있다. 그리고 Hamming Distance란 길이가 같은 두 DNA가 있을 때, 두 DNA의 차이를 표현하는 것이다. 만약에 "AGCAT"와 "GGAAT"는 첫 번째 글자와 세 번째 글자가 다르므로 Hamming Distance는 2이다.

우리가 할 일은 다음과 같다. n 개의 길이가 같은 DNA가 주어져 있을 때(이 DNA를 $a_1a_2a_3a_4...$ 이라고 하자) Hamming Distance의 합이 가장 작은 DNA s 를 구하는 것이다. 즉, s 와 a_1 의 Hamming Distance + s 와 a_2 의 Hamming Distance + s 와 a_3 의 Hamming Distance ... 의 합이 최소가 된다는 의미이다.

입력

첫 줄에 DNA의 수 N 과 문자열의 길이 M 이 주어진다. 그리고 둘째 줄부터 $N+1$ 번째 줄까지 N 개의 DNA가 주어진다. N 은 1,000보다 작거나 같은 자연수이고, M 은 50보다 작거나 같은 자연수이다.

출력

첫째 줄에 Hamming Distance의 합이 가장 작은 DNA 를 출력하고, 둘째 줄에는 그 Hamming Distance의 합을 출력하시오. 그러한 DNA가 여러 개 있을 때에는 사전순으로 가장 앞서는 것을 출력한다.

예제 입력 1 복사

```
5 8
TATGATAC
TAAGCTAC
AAAGATCC
TGAGATAC
TAAGATGT
```

예제 출력 1 복사

```
TAAGATAC
7
```

예제 입력 2 복사

```
4 10
ACGTACGTAC
CCGTACGTAG
GCGTACGTAT
TCGTACGTAA
```

예제 출력 2 복사

```
ACGTACGTAA
6
```

예제 입력 3 복사

```
6 10
ATGTTACCAT
AAGTTACGAT
AACAAAGCAA
AAGTTACCTT
AAGTTACCAA
TACTTACCAA
```

예제 출력 3 복사

```
AAGTTACCAA
12
```

출처

ACM-ICPC (/category/1) > Regionals (/category/7) > Asia (/category/42) > Korea (/category/211) > Asia Regional - Seoul 2006 (/category/detail/1066) B번

- 빠진 조건을 찾은 사람: jsa3824 (/user/jsa3824)
- 잘못된 데이터를 찾은 사람: tncks0121 (/user/tncks0121)

링크

- ACM-ICPC Live Archive (https://icpcarchive.ecs.baylor.edu/index.php?option=com_onlinejudge&Itemid=8&page=show_problem&problem=1603)
- ZJU Online Judge (<http://acm.zju.edu.cn/onlinejudge/showProblem.do?problemCode=3132>)