

# Algorithm Design Laboratory with Applications

Prof. Stefano Leucci

## Problem: *Omicron Persei 8*.

Life has been discovered on the planet Omicron Persei 8. Their genetic material is not made of the familiar adenine (A), cytosine (C), guanine (G), or thymine (T), but rather of Bolonium (B), Dalekaniumm (D), Hadamardium (H), and Unobtanium (U).

In order to reconstruct the phylogenetic trees of these new organisms you need to measure how different two strings  $s, t \in \{B, D, H, U\}^*$  of sequenced genetic material are from each other.

Of course, even if the organisms are very similar, their genetic material will have some gaps and dissimilarities. To measure how far  $s$  is from  $t$  while accounting for these differences, you allow yourself to consider two strings  $s'$  and  $t'$  derived from  $s$  and  $t$  by adding gaps (denoted by a dash -), with the constraint that  $s'$  and  $t'$  must have the same length.

For example, instead of  $s = \text{HBHHUDDDBHDDUBDHUUDDBUH}$  you might consider the string  $s' = \text{HBHHUD-DBHDDUBDHU--UUDDBUH}$ , and instead of  $t = \text{BHHUDUBHDDBDHUDHHUUDDBBH}$  you might consider the string  $t' = \text{-BHHUDUBHDD-BDHUDHHUUDDBBH}$ .

The dissimilarity  $\delta(s', t')$  between  $s'$  and  $t'$  is the number of positions  $i = 1, \dots, |s'|$  for which the  $i$ -th characters of  $s'$  and  $t'$  are not equal. For example,  $\delta(s', t') = 6$  since there are exactly 6 mismatches between  $s'$  and  $t'$ , highlighted in red:

```
HBHHUD-DBHDDUBDHU--UUDDBUH
-BHHUDUBHDD-BDHUDHHUUDDBBH
```

The distance  $\Delta(s, t)$  between  $s$  and  $t$  is the smallest value of  $\delta(s', t')$  attainable for the best choice of  $s'$  and  $t'$ . Your task is to design an algorithm that, given  $s$  and  $t$ , computes  $\Delta(s, t)$ .

**Input.** The input consists of a set of instances, or *test-cases*, of the previous problem. The first line contains the number  $T$  of test-cases. Each test case consists of 2 lines. The first line contains the string  $s$ , while the next line contains the string  $t$ .

**Output.** The output consists of  $T$  lines. The  $i$ -th line is the answer to the  $i$ -th test-case and contains  $\Delta(s, t)$ .

**Assumptions.**  $1 \leq T \leq 10$ ;  $1 \leq |s|, |t| \leq 2^{11}$ .

### Example.

*Input:*

---

```
1
HBHHUDDDBHDDUBDHUUDDBUH
BHHUDUBHDDBDHUDHHUUDDBBH
```

---

*Output:*

---

```
6
```

---

**Requirements.** Your algorithm should require  $O(|s| \cdot |t|)$  time (with reasonable hidden constants).

**Notes.** A reasonable implementation should not require more than 1 second for each input file.