ETH Zurich
Institut für Theoretische Informatik
Prof. Dr. Angelika Steger
Prof. Dr. Emo Welzl
Prof. Dr. Peter Widmayer

# Algorithms Lab

### Exercise – *Evolution*

All life on earth shares a common ancestor known as the last universal common ancestor (LUCA) which lived approximately $3.5$–$3.8$ billion years ago. Repeated formation of new species (speciation), change within species (anagenesis), and loss of species (extinction) throughout the evolutionary history of life on earth are demonstrated by shared sets of morphological and biochemical traits, including shared DNA sequences. These shared traits are more similar among species that share a more recent common ancestor, and can be used to reconstruct a biological 'tree of life' based on evolutionary relationships (phylogenetics), using both existing species and fossils.

In a recent research project a team of scientists reconstructed the tree of life, which is a rooted tree (with LUCA at the root), where the vertices represent species. Each species, except the root, has exactly one neighbour towards the root (its immediate ancestor). Moreover, each species has an arbitrary number of neighbours (possibly zero) in direction away from the root (its immediate offspring). Additionally, each species has an age, this is how many years ago it developed. Of course, the ancestors of a species are older and its offspring are younger than the species itself.

In a follow-up project the scientists want to build a database for the species containing the tree of life and answering queries of the following form. Given a species and an age, find the oldest ancestor of this species which is not older than the specified age. This would help them to answer questions like: 'Were the ancestors of my colleague 3 billion years ago really unicellular fungi¿. Your task is to assist the team and help implementing the database.

**Input**    The first line of the input contains the number $t \leq 50$ of test cases. Each of the $t$ test cases is described as follows.

- It starts with a line that contains two integers n q, separated by a space, denoting the number of species ($1 \leq n \leq 50000$) and the number of queries ($1 \leq q \leq 50000$), respectively.

- The following $n$ lines define the species. Each line contains the name s of the species and its age a separated by space. A name is a string over the alphabet $\{'a', \ldots, 'z'\}$ consisting of at least one and at most 10 characters and the age is an integer with $0 \leq a \leq 10^9$.

- The following $n-1$ lines define the 'is immediate offspring of' relation of the species. Each line contains two names s and p of species separated by space indicating that $s$ is an immediate offspring of $p$.

- The following $q$ lines define the queries. Each line contains the name s of the species to query and the maximum age b with $0 \leq b \leq 10^9$. Moreover, you can assume that $b$ is at least the age of $s$.

**Output**    For each test case the corresponding output appears on a separate line. Output for each query $s$ and $b$ the name of the oldest species which is an ancestor of $s$ (i.e. it appears on the unique path from $s$ to the root in the tree of life) and which is not more than $b$ years old, separated by space, in the same order as the queries were read in.

**Points**   There are two groups of test sets, worth 100 points in total.

1. For the first group of test sets, worth 30 points, you may assume that $n, q \leq 1000$.

2. For the second group of test sets, worth 70 points, there are no additional assumptions.

Corresponding sample test sets are contained in `testi.in/out`, for $i \in \{1, 2\}$.

**Sample Input**

```
1
8 3
luca 10
bacteria 7
archaea 8
eukaryota 7
aquifex 3
plants 4
animals 6
thermotoga 1
bacteria luca
archaea luca
eukaryota luca
aquifex bacteria
plants eukaryota
animals eukaryota
thermotoga aquifex
thermotoga 4
thermotoga 2
plants 8
```

**Sample Output**

```
aquifex thermotoga eukaryota
```