

Practical 4: NCBI Taxonomy

Name Surname

Name Surname

10/10/2023, submission deadline 16/10/2023

Solve the following exercise in groups of two students. Write the Python scripts, perform the computations, and make the graphics that are asked for (if any) in the practical below. Write your solution in a \LaTeX document and generate a PDF file with your solution. Take care to number your answers exactly as in this exercise. Upload your solution in PDF format to the web page of the course at `raco.fib.upc.edu` no later than the submission deadline.

You can make use of the Python package **networkx** (and other packages) to compute your answers, as you please. The datasets (if any) can be downloaded from the web page of the course at `raco.fib.upc.edu`.

1. (40 points) Given a file **nodes.dmp** for the NCBI taxonomy, write a Python script to store it in a rooted tree. Give the code of your Python script as your answer to this question, using the \LaTeX package **listings**.
2. (5 points) Is the NCBI taxonomy a rooted tree, a directed acyclic graph, or a directed graph with cycles?
3. (5 points) How many nodes are there in the NCBI taxonomy?
4. (40 points) Write a Python script to restrict the rooted tree to the seven standard taxonomic ranks: *kingdom*, *phylum*, *class*, *order*, *family*, *genus*, *species*. Give the code of your Python script as your answer to this question, using the \LaTeX package **listings**.
5. (5 points) What is the name of the *kingdom* taxonomic rank in the NCBI taxonomy?
6. (5 points) How many nodes are there in the NCBI taxonomy, once restricted to the seven standard taxonomic ranks?

```
\documentclass[12pt,a4paper]{article}
\usepackage{listings}
\usepackage{mathptmx}
\usepackage{savetrees}
\title{Practical 4: NCBI Taxonomy}
\author{Name Surname \and Name Surname}
\date{10/10/2023, submission deadline 16/10/2023}
\begin{document}
\maketitle
\begin{enumerate}
\item ...
\item ...
\item ...
\item ...
\item ...
\item ...
\end{enumerate}
\end{document}
```