



IIC2115 – Programación como Herramienta para la Ingeniería (II/2022)

## Laboratorio 1

### Aspectos generales

- **Objetivo:** evaluar individualmente el aprendizaje sobre el uso de técnicas de POO y estructuras de datos para manejar árboles filogenéticos.
- **Lugar de entrega:** jueves 8 de septiembre a las 23:59 hrs. en repositorio privado.
- **Formato de entrega:** archivo Python Notebook (**L1.ipynb**) con la solución del laboratorio. El archivo debe estar ubicado en la carpeta **L1**. Utilice múltiples celdas de texto y código para facilitar el trabajo del cuerpo docente. Laboratorios que no cumplan el formato de entrega tendrán un descuento de 0,5 pts a su puntaje final.
- **Entregas atrasadas:** El descuento por atraso se realizará de acuerdo a lo definido en el programa del curso. Si su laboratorio es entregado fuera de plazo, tiene hasta el **viernes 9 de septiembre a las 11:59 AM** para responder el formulario de **entregas fuera de plazo** disponible en el Syllabus.
- **Issues:** Las discusiones en las *issues* del Syllabus que sean relevantes para el desarrollo del laboratorio, serán destacadas y se considerarán como parte de este enunciado. Así mismo, el uso de librerías externas que solucionen aspectos fundamental del laboratorio no podrán ser utilizadas. Solo se podrán utilizar las que han sido aprobadas en las *issues* y el enunciado.
- **Laboratorios con errores de sintaxis y/o que generen excepciones en todas las ejecuciones** serán calificados con **nota 1.0**.

## El problema: Desorden evolutivo

Se han acercado a usted miembros de un grupo de investigación de la facultad de Ciencias Biológicas. Están estudiando las relaciones evolutivas de varios organismos pero tienen tantos ejemplares que ya se ha vuelto un desafío armar las relaciones evolutivas entre especies a mano. El grupo necesita una solución automatizada y piden su ayuda. Su objetivo es ayudar a los investigadores a construir árboles filogenéticos a partir de un grupo de especies de interés y poder responder un conjunto de consultas sobre este árbol.

## Árbol Filogenético

Los árboles filogenéticos son diagramas que representan las relaciones evolutivas entre organismos o especies<sup>1</sup> de acuerdo a un grupo de características observables. Un ejemplo real es el que se ve en la figura 1

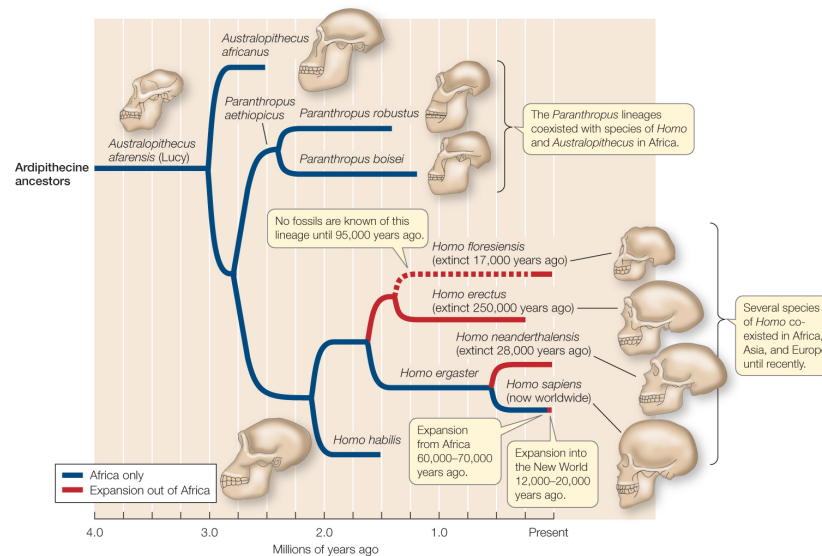
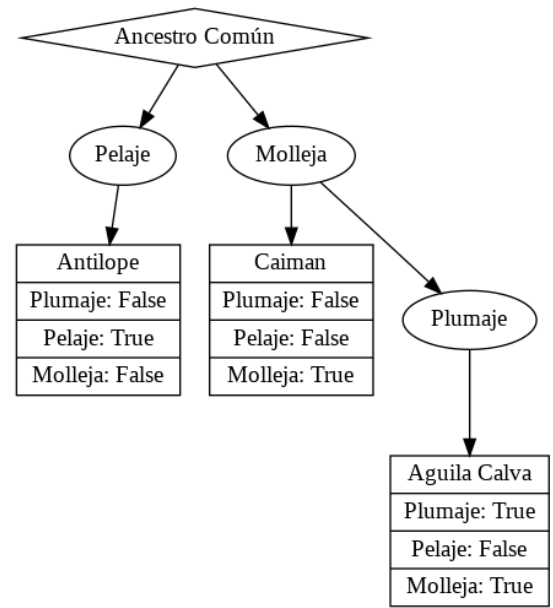


Figura 1: Árbol Filogenético Homínidos

Este tipo de árboles se generan a partir de una tabla que contiene las especies de interés e información de la presencia de ciertas características relevantes. Un ejemplo de lo que busca el grupo de investigación se puede ver a continuación.

<sup>1</sup>Fuente: Khan Academy - Curso Biología Avanzada, Unidad: Selección Natural, Capítulo: Filogenia

	Plumaje	Molleja	Pelaje
<b>Antílope</b>	-	-	✓
<b>Caimán</b>	-	✓	-
<b>Águila calva</b>	✓	✓	-



Cuadro 1: este ejemplo muestra a la izquierda una lista de especies de interés y las características que cumple cada una. Al lado derecho, muestra el árbol resultante.

Como se puede apreciar, los árboles filogenéticos tipo tienen tres de nodos:

1. Los que representan a las especies de interés y listan sus características. Tienen forma rectangular.
2. Los que representan el ancestro común y representa la raíz del árbol. Tiene forma de diamante.
3. Los nodos que indican características evolutivas, como 'Pelaje' y 'Plumaje'. Tienen forma ovalada.

Este último tipo está representando una especie desconocida que desarrolló la característica que indica su etiqueta. O sea, hay un ancestro del caimán y el águila calva que desarrolló 'Molleja'. A partir de este, evolucionó el caimán y, al mismo tiempo, generó la rama evolutiva que lleva hasta la águila calva.

## Formalización

Considerando lo que sabe, ha decidido dividir este problema en tres etapas:

- Leer datos y crear las clases necesarias para armar el árbol y todos sus tipos de nodos (2.0 pts.)
- Usar los datos para armar el árbol correspondiente (2.0 pts.)
- Generar métodos en la clase árbol que respondan a las posibles consultas (2.0 pts.)

**Hint:** no olvide revisar los Notebooks de ejemplo, especialmente "05 - Sets.ipynb" para implementación y uso de conjuntos, y "06 - Árboles.ipynb" para la implementación de las clases `Árbol` y `Nodo`, y los algoritmos BFS y DFS para recorrerlo.

## Leer datos y crear las clases necesarias (2.0 puntos)

Los datos se entregarán desde un archivo; el camino al archivo debe entregar a su programa usando el comando `input()`. Los datos tendrán el siguiente formato:

- La primera línea contiene dos números enteros: la cantidad de características a revisar  $C$  y la cantidad de especies de interés  $N$ .
- La siguiente línea contiene los nombres de las características, separados por comas.
- Las siguientes  $N$  líneas contienen el nombre de la especie y valores `True` o `False` para cada característica, separado por comas.

Un ejemplo:

```
1 3 3
2 plumaje,molleja,pelaje
3 antilope,false,false,true
4 caiman,false,true,false
5 aguila calva,true,true,false
```

Se recomienda guardar esta información en alguna estructura de datos intermedia antes de usarla para crear un árbol filogenético. Se debe imprimir esta estructura.

Además, debe generar una clase `Arbol`, una clase abstracta `Nodo` y una clase para cada tipo de nodo. La clase `Nodo` debe tener información sobre su nodo progenitor, información de sus nodos hijos y un método abstracto que imprima su información. Las clases de cada tipo de nodo deben tener una implementación propia de este método. Fuera de eso, tiene libertad de generar los atributos y métodos que considere necesarios.

Incluya ejemplos simples y propios que demuestren el funcionamiento de sus clases.

## Armar un árbol filogenético de los datos (2.0 puntos)

La idea detrás de armar un árbol filogenético, se basa en la búsqueda de características comunes. Si varias especies tienen una característica común, todas ellas se consideran sucesoras de la especie desconocida que

desarrolló este atributo. Por ejemplo, en la figura del inicio, tanto el caimán como el águila calva comparten la característica de 'Molleja', pero el antílope, no. Por lo tanto, el caimán y el águila calva deben ser sucesores de 'Molleja'.

Su deber es pensar un algoritmo que haga esta transformación de los datos a un árbol filogenético. Debe explicar su funcionamiento paso a paso en bloques de texto de su Notebook e implementar su funcionamiento.

Para guiar este análisis, considere que:

- Siempre habrá un ancestro inicial en representación de la raíz del árbol.
- Las especies siempre son nodos sin hijos; son el último nodo de una rama evolutiva.
- Existen especies que pueden no tener ninguna de las características de la lista; estas se conocen como elementos externos. Los elementos externos heredan del ancestro común que representa el nodo raíz. Ver sección de ejemplos adicionales, ejemplo 1.
- La jerarquía de características se decide de acuerdo a conjuntos. Ejemplo: en la figura del inicio, la característica 'Molleja' tiene dos especies: el caimán y el águila calva, mientras que 'Plumaje' sólo tiene al águila. Si los vemos como conjuntos, las especies de 'Plumaje' son un subconjunto de 'Molleja', por lo que 'Molleja' debe estar más arriba en la jerarquía.
- Es posible tener situaciones en las que una característica tenga más de dos hijos; esto se conoce como politomía. Ver sección de ejemplos adicionales, ejemplo 2.

## **Lista de consultas (2.0 puntos)**

Las consultas que se piden:

- Dada una especie, imprimir su rama evolutiva comenzando desde la raíz.
- Encontrar la rama evolutiva más larga e imprimirla. Si hay más de una, imprimir sólo una de ellas.
- Encontrar la rama evolutiva más corta e imprimirla. Si hay más de una, imprimir sólo una de ellas.
- Dadas dos especies, imprimir el primer ancestro común.
- Imprimir el árbol completo en consola; imprimir cada nivel del árbol en una nueva línea.

## Bonus

Se puede obtener un bono de 5 décimas sobre la nota del laboratorio, si se implementa una visualización del árbol en formato similar a los ejemplos mostrados antes. Se pueden usar librerías como GraphViz. Para optar a esta bonificación, la nota del laboratorio debe ser mayor o igual a 3,5.

# Ejemplos adicionales

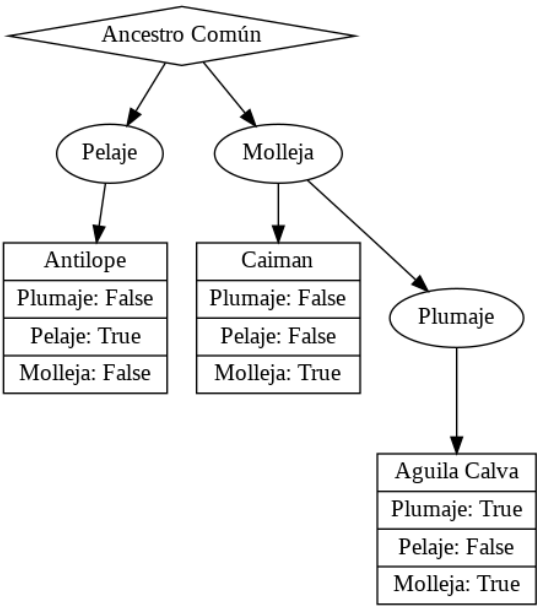
## Ejemplo 0: sin elemento externo

Archivo:

```
1 3 3
2 plumaje,molleja,pelaje
3 antilope,false,false,true
4 caiman,false,true,false
5 aguila calva,true,true,false
```

Tabla y Árbol

	Plumaje	Molleja	Pelaje
Antílope	-	-	✓
Caimán	-	✓	-
Águila calva	✓	✓	-



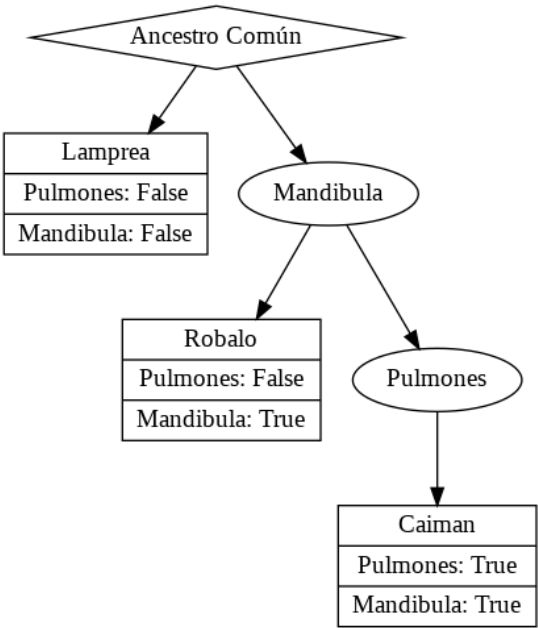
Ejemplo 1: elemento externo

Archivo:

```
1 2 3
2 mandibula ,pulmones
3 lamprea ,false ,false
4 robalo ,true ,false
5 caiman ,true ,true
```

Tabla y Árbol

	Mandíbula	Pulmones
Lamprea	-	-
Róbalo	✓	-
Caimán	✓	✓





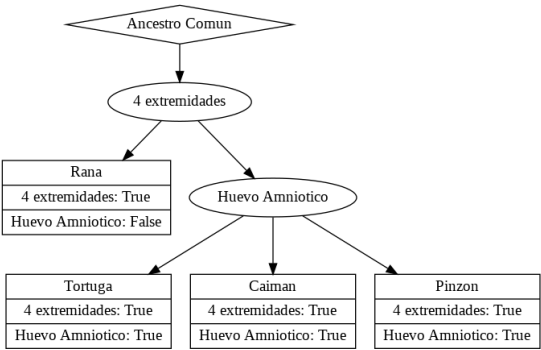
## Ejemplo 2: politomía

Archivo:

```
1 2 4
2 4 extremidades,huevo amniotico
3 rana,true,false
4 tortuga,true,true
5 caiman,true,true
6 pizon,true,true
```

Tabla y Árbol

	4 Extremidades	Huevo Amniótico
Rana	✓	-
Tortuga	✓	✓
Caimán	✓	✓
Pinzón	✓	✓



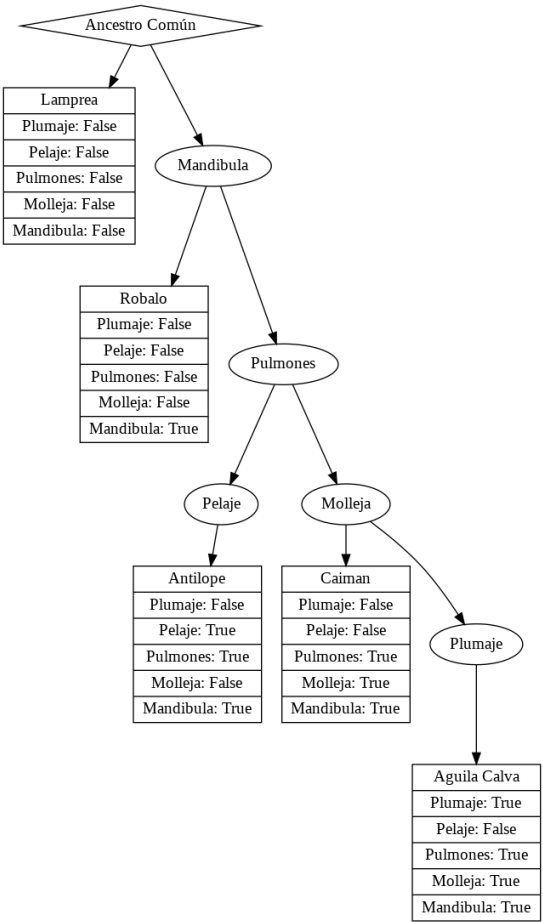
### Ejemplo 3: árbol grande

Archivo:

```
1 5 5
2 plumaje ,pelaje ,pulmones ,molleja ,mandibula
3 lamprea ,false ,false ,false ,false ,false
4 antelope ,false ,true ,true ,false ,true
5 robalo ,false ,false ,false ,false ,true
6 aguila calva ,true ,false ,true ,true ,true
7 caiman ,false ,false ,true ,true ,true
```

Tabla y Árbol

	Plumaje	Pelaje	Pulmones	Molleja	Mandíbula
Lamprea	-	-	-	-	-
Antílope	-	✓	✓	-	✓
Róbalo	-	-	-	-	✓
Águila Calva	✓	-	✓	✓	✓
Caimán	-	-	✓	✓	✓



## Política de Integridad Académica

Los alumnos de la Escuela de Ingeniería deben mantener un comportamiento acorde al Código de Honor de la Universidad:

*“Como miembro de la comunidad de la Pontificia Universidad Católica de Chile me comprometo a respetar los principios y normativas que la rigen. Asimismo, prometo actuar con rectitud y honestidad en las relaciones con los demás integrantes de la comunidad y en la realización de todo trabajo, particularmente en aquellas actividades vinculadas a la docencia, el aprendizaje y la creación, difusión y transferencia del conocimiento. Además, velaré por la integridad de las personas y cuidaré los bienes de la Universidad.”*

En particular, se espera que mantengan altos estándares de honestidad académica. Cualquier acto deshonesto o fraude académico está prohibido; los alumnos que incurran en este tipo de acciones se exponen a un procedimiento sumario. Ejemplos de actos deshonestos son la copia, el uso de material o equipos no permitidos en las evaluaciones, el plagio, o la falsificación de identidad, entre otros. Específicamente, para los cursos del Departamento de Ciencia de la Computación, rige obligatoriamente la siguiente política de integridad académica en relación a copia y plagio: Todo trabajo presentado por un alumno (grupo) para los efectos de la evaluación de un curso debe ser hecho individualmente por el alumno (grupo), sin apoyo en material de terceros. Si un alumno (grupo) copia un trabajo, se le calificará con nota 1.0 en dicha evaluación y dependiendo de la gravedad de sus acciones podrá tener un 1.0 en todo ese ítem de evaluaciones o un 1.1 en el curso. Además, los antecedentes serán enviados a la Dirección de Docencia de la Escuela de Ingeniería para evaluar posteriores sanciones en conjunto con la Universidad, las que pueden incluir un procedimiento sumario. Por “copia” o “plagio” se entiende incluir en el trabajo presentado como propio, partes desarrolladas por otra persona. Está permitido usar material disponible públicamente, por ejemplo, libros o contenidos tomados de Internet, siempre y cuando se incluya la cita correspondiente.