Evolutionary Trees

- Although hierarchical classification is implicit in Linnaeus’s taxonomy, his writings use tables rather than branching trees to categorize each species.
- The modern conception in which living things form a tree that splits as new species appear came into prominence after Charles Darwin published *The Origin of Species* in 1859.
- The branching diagram on the right appears in Darwin’s notebooks and offers his first example of a tree as a model for evolution.

Trees

- A tree is a collection of objects called nodes that begins with a single root node that is connected to other nodes in a branching pattern that includes no cycles.
- Trees appear in many familiar contexts including, for example, family trees and the evolutionary tree developed by Darwin.
- Trees also form the basis for the object-oriented hierarchies used in languages like Python.

Family Trees

- One of the most common examples of a tree is the family tree, illustrated here by the House of Normandy:

  - This example is useful for defining terminology:
    - William I is the root of the tree.
    - Adela is a child of William I and the parent of Stephen.
    - Robert, William II, Adela, and Henry I are siblings.
    - Henry II is a descendant of William I, Henry I, and Matilda.
    - William I is an ancestor of everyone else in this tree.

Binary Search Trees

- The tree that supports the implementation of a dictionary is called a binary search tree (or BST for short). Each node in a BST has exactly two subtrees: a left subtree that contains all the nodes that come before the current node and a right subtree that contains all the nodes that come after it. Either or both of these subtrees may be None.
- The classic example of a binary search tree uses the names from Walt Disney’s *Snow White and the Seven Dwarves*:

A Question of Balance

- Ideally, a binary search tree containing the names of Disney’s seven dwarves looks like this:

  - If, however, you happened to enter the names in alphabetical order, this tree would end up being a simple linked list in which all the left subtrees were None and the right links formed a simple chain. Algorithms on that tree would run in \(O(N)\) time instead of \(O(\log N)\) time.
- A binary search tree is balanced if the height of its left and right subtrees differ by at most one and if both of those subtrees are themselves balanced.
Exercise: Balanced Trees

• Which of the following binary search trees are balanced?

(a) Happy  Dopey  Sneezy  Bashful  Grumpy  Doc  Sleepy
(b) Grumpy  Dopey  Sneezy  Bashful  Happy  Doc  Sleepy
(c) Sleepy  Grumpy  Dopey  Bashful  Happy  Doc  Sneezy
(d) Sleepy  Happy  Sneezy  Bashful  Grumpy  Doc  Sneezy

AVL Trees

• The easiest tree-balancing algorithm to describe is the earliest one. This approach is called AVL trees after the initials of the Russian mathematicians Georgy Adelson-Velsky and Evgenii Landis who published the technique in 1962.

• The central idea behind the AVL algorithm is to keep track of a balance factor in each node that tracks the relative heights of its subtrees. In the diagrams that follow, the balance factor appears as a symbol in the upper right corner of the node.

• The algorithm is happy if the subtrees are the same height (noted by the symbol =) or differ by one in some direction (the symbols + and -). If adding a node causes the balance factor to become ++ or --, the tree must be rebalanced.

• The AVL algorithm ensures that balance can be restored using either one or two rotations, as shown on the next few slides.
Illustrating the AVL Algorithm

H
He
Li
Be
B
C

Illustrating the AVL Algorithm

H
He
Li
Be
B
C

Illustrating the AVL Algorithm

H
He
Li
Be
B
C

Illustrating the AVL Algorithm

H
He
Li
Be
B
C

Illustrating the AVL Algorithm

H
He
Li
Be
B
C

Illustrating the AVL Algorithm

H
He
Li
Be
B
C
Recursive Structure of Expressions

- In most programming languages, an expression is a recursive structure that can contain subexpressions.
- Expressions in a simple language might have the following definition, in which the last two cases are recursive:
  - An integer constant
  - A variable name that holds an integer value
  - Two expressions joined by an operator
  - An expression enclosed in parentheses
- Before the Python interpreter can work with expressions, it must convert the string representation into a recursive data structure that mirrors the recursive definition of expressions. This process is called parsing.
Parsing and Grammars

- Interestingly, the theoretical foundations of parsing used in modern programming languages does not come from computer science but rather from linguistics.
- The linguist and political activist Noam Chomsky formalized the theory of parsing in 1956 by describing a hierarchy of grammatical structures that he used to describe human languages. Programming languages typically use the level of the Chomsky hierarchy called context-free grammars, in which each syntactic unit can be replaced according to a set of rules independent of the surrounding context.
- Programming language translators like the Python interpreter define a grammar for the language and then use that grammar to create a parse tree that shows how the text of a program follows the grammatical rules. This process is illustrated on the next slide.

Creating a Parse Tree

\[
x = 2 \times (x + 4 \times y)
\]