Data Structures and Algorithms with Python

Heikki Peura

h.peura@imperial.ac.uk

Lecture 3

Last time

Algorithm complexity

- Asymptotic analysis and Big-Oh notation
- Searching and sorting
- Divide-and-conquer

Plan for today:

- Introduction to data structures
- Object-oriented programming (OOP)

Python data types?

Data types in Python:

- `int, float, str, None, list, tuple, set, dict`

When are different data structures appropriate?

- Compound data types: `list, tuple, set, dict`?
- Ordered sequences: `list, tuple`
- Unordered: `set, dict`

`list` VS `dict`

Suppose you're designing a game where players catch **pocket monsters** and make them fight each other

- You need to keep track of the monsters and their attributes

```
# Using lists?
monsters = ['Pikachu','Squirtle','Mew']
combatPoints = [20,82,194]

# Or like this?
monsters = [['Pikachu',20],['Squirtle',82],['Mew',194]]

# But how to get Pikachu by its name?
# O(log n) to search for name even if sorted
```

```
# Using a dictionary?
monsters = {'Pikachu':20,'Squirtle':82,'Mew':194}
monsters['Pikachu']
# Turns out to be O(1) operation for this data structure (hash table magic!)
```

Data structures

How to organize data for quick access?

- Like with algorithms: recipe $\rightarrow$ translate to Python

Examples: lists, stacks, queues, dictionaries (hash tables), graphs, trees, etc

Different data structures are suitable for different tasks

- Support different sets of operations (`list` vs `dict`)
- How to choose?

Levels of life

1. “What’s a data structure?”
2. Comfortable at cocktail parties
3. “Guys, we need to use a dictionary for this problem”
4. “I only use data structures that I create myself”

We have already been using data structures

```
'Hello World'
3.14159
9
L=[1,1999,0,-2,9]
```

These are all **objects**. An object has:

- A type: `int`, `str`, `list` (`L` is an **instance** of a `list`)
- An internal representation of data
- A set of functions that operate on that data

A Python list has many operations

Some of the operations:

- `len(L)`,`max(L)`,`min(L)`,...
- `L[start:stop:step]`: returns elements of `L` from `start` to `stop` with step size `step`
- `L[i]= e`: sets the value at index `i` to `e`
- `L.append(e)`: adds `e` to the end of `L`
- `L.count(e)`: returns how many times `e` occurs in `L`
- `L.insert(i,e)`: inserts `e` at index `i` of `L`
- `L.extend(L1)`: appends the items of `L1` to the end of `L`
- `L.remove(e)`: deletes the first occurrence of `e` from `L`
- `L.index(e)`: returns the index of first occurrence of `e` in `L`
- `L.pop(i)`: removes and returns the item at index `i`, default `i = -1`
- `L.sort()`: sorts elements of `L`
- `L.reverse()`: reverses the order of elements of `L`

A list is an `object`

An **object** has:

- A type: `int`, `str`, `list` (`L` is an **instance** of a `list`)
- An internal representation of data
- A set of functions that operate on that data

```
L = [1,1999,0,-2,9]
L.append(8)
L.insert(2,1000)
t = L.pop()
L.remove(1)
help(L)
```

The point:

- Interface: the user knows what she can do with a list
- Abstraction barrier: the user does not need to know the details of what goes on under the hood (similarly to functions)
- Invaluable in managing complexity of programs

List operations complexity?

```
L = [1,1999,0,-2,9]
L.append(9)
t = L[2]
t = L.pop(0)
```

We have **assumed** that list operations like retrieving or adding an item are $O(1)$

- A list has internal functions with algorithms to perform these operations

But we have seen that **how we write algorithms matters** a lot...

- Does it matter how you implement a list?
- Yes!
- What is the internal data representation of a list?

How can we implement a list?

For a list, we would like to:

- Add and remove elements, look up values, change values, . . .

An indexed **array**?

56	24	99	32	9	61	57	79

- Reserve n slots of memory for list from computer memory
- Easy to access an item at index `i`: $O(1)$
- Easy to add an item to end: $O(1)$ (but details are advanced...)
- **Difficult to add item to beginning**: $O(n)$ (need to move all other elements)

A linked list of nodes?

Linked list?

$\boxed{56} \rightarrow \boxed{24} \rightarrow \boxed{99} \rightarrow \boxed{32} \rightarrow \boxed{9} \rightarrow \boxed{61} \rightarrow \boxed{57} \rightarrow \boxed{79}$

- **Each node contains information on the next node** – the list itself just knows the first and last one
- Easy to add items to either end: $O(1)$
- **Difficult to look up item by index**... $O(n)$ (need to walk through the nodes)

A Python list has many operations

Some of the operations:

- `len(L)`, `max(L)`, `min(L)`, ...
- `L[start:stop:step]`: returns elements of `L` from `start` to `stop` with step size `step`
- `L[i]= e`: sets the value at index `i` to `e`
- `L.append(e)`: adds `e` to the end of `L`
- `L.count(e)`: returns how many times `e` occurs in `L`
- `L.insert(i,e)`: inserts `e` at index `i` of `L`
- `L.extend(L1)`: appends the items of `L1` to the end of `L`
- `L.remove(e)`: deletes the first occurrence of `e` from `L`
- `L.index(e)`: returns the index of first occurrence of `e` in `L`
- `L.pop(i)`: removes and returns the item at index `i`, default `i = -1`
- `L.sort()`: sorts elements of `L`
- `L.reverse()`: reverses the order of elements of `L`

Object-oriented programming (OOP)

Everything is an object with a type: `L=[1,2,3,4]` is an **instance** of a `list` object

Abstraction — creating an object type:

- Define internal representation and interface for interacting with object — user only needs interface
- Then we can create new **instances** of objects and delete them

This is **"divide-and-conquer" development**

- Modularity — treat a complex thing like a list as primitive
- Easier to reuse code — keep code clean: eg '+' method for integers and strings

Why define objects?

Suppose you're designing a game where players catch **pocket monsters** and make them fight each other

```python
# Using lists?
monsters = ['Pikachu','Squirtle','Mew']
combatPoints = [20,82,194]
hitPoints = [53,90,289]
```

```python
# Using a dictionary?
monsters = {'Pikachu':[20,53],'Squirtle':[82,90],'Mew':[194,289]}
```

Defining an object type

An object contains

- **Data**: attributes (of a monster)
- **Functions**: methods that operate on that data

Suppose you're designing a game where players catch **pocket monsters** and make them fight each other

```python
class Monster():
 """ Attributes and methods
        """
```

`class` statement defines new object type

Creating a new monster

Attributes of a monster?

```
class Monster():
    def __init__(self,combatPoints):
        self.combatPoints = combatPoints

Pikachu = Monster(65)
Squirtle = Monster(278)
print(Pikachu.combatPoints)
```

`self`: Python passes the object itself as the first argument — convention to use word "self"

- But you omit this when calling the function

- Notice the "`.`" operator (like with a `list`)
- The `__init__` method is called when you call `Monster()`

Growing our monsters

```python
class Monster():
    def __init__(self,name,combatPoints,hitPoints):
        self.name = name
        self.combatPoints = combatPoints
        self.hitPoints = hitPoints
        self.health = hitPoints

    def hurt(self,damage):
        self.health = self.health - damage
        if self.health <= 0:
            print(self.name + ' is dead!')
```

OOP is wonderful

Easy to handle many “things” with **common attributes**

Abstraction **isolates** the use of objects from implementation details

Build **layers of abstractions** — our own on top of Python’s classes

- Keeping track of different monsters and their attributes

Accessing data structures

```
class Monster():
    def __init__(self,name,combatPoints,hitPoints):
        self.name = name
        self.combatPoints = combatPoints
        self.hitPoints = hitPoints
        self.health = hitPoints

    def getCombatPoints(self): # access data through function
        return self.combatPoints

    # more code in the class...

cp = Pikachu.combatPoints # not great - what if you make a mistake?
cp = Pikachu.getCombatPoints() # much better - cannot mess stuff up
```

Workshop: data structures and OOP

After the break...

More monsters

Implementing data structures: queues