Designing Programs with Inheritance

Welcome back to CS 2100!

Prof. Rasika Bhalerao

Poll: What would be a good relationship between `Chair` and `Throne`?

1. `Throne` should be an interface implemented by `Chair`
2. `Chair` should be an interface implemented by `Throne`
3. `Throne` should be a concrete subclass of the abstract class `Chair`
4. `Chair` should be a concrete subclass of the abstract class `Throne`
5. None of the above

Inheritance vs composition

Inheritance: *is a* relationship

- A square *is a* rectangle.
- One of the classes is a subclass of the other.
- (One of the classes may be abstract, but neither should be an interface.)

Composition: *has a* relationship

- A square *has* four edges.
- One of the classes holds an instance of the other class as an instance variable.

Good object-oriented design requires knowing when to use inheritance versus composition.

Example when to use composition (*has a*)

`SocialMedia` : a class that holds information about a social media platform, including a set of users

- Correct: composition. `SocialMedia` class has a `set[User]` as an attribute
- Incorrect: inheritance. It would be wrong to make the `SocialMedia` class extend the `set[User]` class

  Both are possible to do using Python, but the inheritance version is silly:

```
class SocialMedia(set[str]):
    pass

fb = SocialMedia()
fb.add('Mini')
fb.add('Binnie')

print(fb)  # SocialMedia({'Mini', 'Binnie'})
```

Example when to use composition (*has a*)

A `House` with a kitchen and bedroom

- Correct: composition. The `House` class should have instance variables for a `Kitchen` and a `Bedroom`
- Incorrect: inheritance. It would be wrong to make the `House` class extend the `Kitchen` class and add the features of a `Bedroom` (like a `Bed` instance variable)

Again, both options are possible to do using Python, but the inheritance version would require admitting that one's house is a specific type of kitchen.

Example when to use inheritance (*is a*)

`Cat`, `Lion`, and `HouseCat`:

- Correct: inheritance. The `Lion` and `HouseCat` classes should extend the `Cat` class
- Incorrect: composition. It would be wrong to make the `Lion` and `HouseCat` classes each have an instance variable for a `Cat` to which they outsource all of the kneading

Poll: Which of these pairs of classes should use inheritance rather than composition?

1. `VideoGame` and `Physics`
2. `UIComponent` and `TextBox`
3. `Student` and `TA`
4. `OnlineStore` and `Inventory`
5. `TextEditor` and `SpellChecker`

```
from abc import ABC, abstractmethod

class Pet(ABC):
    @abstractmethod
    def express_affection(self) -> None:
        pass

class Cat(Pet):
    def express_affection(self) -> None:
        self.make_biscuits()

    def make_biscuits(self) -> None:
        print('Making biscuits')

class Dog(Pet):
    def express_affection(self) -> None:
        self.slobber()

    def slobber(self) -> None:
        print('Slobbering')

for pet in [Cat(), Dog(), Cat()]:
    pet.express_affection()
```

Polymorphism

This works because of *polymorphism*: the `pet` variable's ability to be both a `Cat` and a `Dog`, and for it to be treated correctly as an instance of both a `Cat` and a `Dog`.

Polymorphism

Let's create classes for `Car`, `Motorcycle`, and `Truck`.

Let's write a function that takes a fleet of vehicles as a list and returns the total fuel needed for the trip.

```
class Vehicle(ABC):
    def __init__(self, mpg: int):
        self.fuel_used: float = 0.0
        self.mpg = mpg

    def move(self, distance: int) -> None:
        self.fuel_used += (distance / self.mpg)

    def get_fuel(self) -> float:
        return self.fuel_used

class Car(Vehicle):
    def __init__(self) -> None:
        super().__init__(26)

class Motorcycle(Vehicle):
    def __init__(self) -> None:
        super().__init__(55)

class Truck(Vehicle):
    def __init__(self) -> None:
        super().__init__(7)
```

```
def get_total_gas(fleet: list[Vehicle]) -> float:
    return sum(veh.get_fuel() for veh in fleet)

fleet: list[Vehicle] = [
    Car(),
    Car(),
    Truck(),
    Motorcycle(),
    Motorcycle(),
    Motorcycle(),
    Motorcycle()
]

for veh in fleet: veh.move(10)

print(get_total_gas(fleet))
```

Design principle: Encapsulation

- Group attributes and methods into a single class.
- Information hiding: discourage direct access to some methods / attributes (using underscores).
    - Protects internal data from unauthorized modification
    - Promotes "modularity" by hiding unnecessary implementation details behind a simple public interface
    - Gives us more flexibility to change implementations without telling the client

Poll: Which class design best demonstrates encapsulation?

1. Expose all internal attributes to the public for flexibility
2. Create a minimal public interface with all complex logic kept private
3. Rather than having a class be a direct subclass of its interface, make it a subclass of a subclass, to add layers of privacy
4. Document internal methods thoroughly for users

Poll: Here is a poorly designed Python class:

```
class Rectangle:
    def __init__(self, width: int, height: int):
        self.width = width
        self.height = height
        self.area = width * height
```

How can we improve its encapsulation?

1. Validate in `__init__()` that `width` and `height` are not negative
2. Make all three attributes private with corresponding getter and setter methods using the `@property` decorator
3. Make `width` and `height` private with corresponding getter/setter `@property` methods, and make `area` a property only (calculated in a getter method)
4. Add docstrings to explain the attributes

Poll: Here is a poorly designed Python class:

```
class Rectangle:
    def __init__(self, width: int, height: int):
        self.width = width
        self.height = height
        self.area = width * height
```

How should errors be handled in an encapsulated class design?

1. Raise all errors to the caller
2. Wrap all errors in `try` / `except` and return error codes instead
3. Wrap low-level internal errors in `try` / `except` and raise them as domain-specific errors
4. Log errors internally but never raise them

Design principle: Coupling vs cohesion

Cohesion: how closely related the parts of a unit are

- closely related to the Single Responsibility Principle
- Example where the unit is a function: aim for it to have a single, well-defined job
- Example where the unit is a class: aim for its methods to be very closely related

Coupling: how dependent different units are on each other

- want to **avoid** this
- Often, that means that one class is too dependent on another, and any changes to the other class will result in "ripple effects" on it.

Bad email sender with too much coupling of its tasks:

```
class BadEmailSender:
    def send_email(self,
                   user: str, email_type_flag: int
    ) -> None:
        if email_type_flag == 1:
            # send a welcome email
        elif email_type_flag == 2:
            # send a password reset email
```

- Many tasks, all very dependent on each other
- Code will be repeated between branches

Better design that uses polymorphism to separate out the tasks into cohesive classes:

```
class Template(ABC):
    @abstractmethod
    def generate_content(self, user: str) -> str:
        pass

class EmailSender:
    def send_email(self,
                   email_template: Template, user: str
    ) -> str:
        return email_template.generate_content(user)

class WelcomeEmail(Template):
    def generate_content(self, user: str) -> str:
        return f"Welcome {user}!"

class PasswordResetEmail(Template):
    def generate_content(self, user: str) -> str:
        return f"Reset password for {user}"
```

Hashing

Poll: What does this print?

```
class Course:
    def __init__(self, department: str, course: int):
        self.department = department
        self.course = course

course_oakland = Course('CS', 2100)
course_boston = Course('CS', 2100)

print(course_oakland == course_boston)
```

1. True
2. False

Poll: How about now?

```
class Course:
    def __init__(self, department: str, course: int):
        self.department = department
        self.course = course

    def __eq__(self, other: object) -> bool:
        if not isinstance(other, Course):
            raise NotImplementedError
        else:
            return self.department == other.department and self.course == other.course


course_oakland = Course('CS', 2100)
course_boston = Course('CS', 2100)

print(course_oakland == course_boston)
```

1. True
2. False

Moving on to the "hashing" topic...

If we try to add a `Course` to a `set`, it raises a `TypeError`

```
class Course:
    def __init__(self, department: str, course: int):
        self.department = department
        self.course = course

    def __eq__(self, other: object) -> bool:
        if not isinstance(other, Course):
            raise NotImplementedError
        else:
            return self.department == other.department and self.course == other.course


course_oakland = Course('CS', 2100)

courses: set[Course] = {course_oakland}  # TypeError: unhashable type: 'Course'
```

And same thing if you try to add it as a key to a `dict`

In order to put someting in a `set` or `dict`, it needs to follow the Hashable protocol

Hashable protocol's required method: `__hash__(self) -> int`

Corresponding interface: `from collections.abc import Hashable`

Why hashing?

If we avoid the `TypeError: unhashable type: 'Course'`:

- Just use a `list` instead
- But `list`s are slower than `set`s in some situations:

```
T = TypeVar('T')

def list_contains(item: T, list: list[T]) -> bool:
    """Returns True if the item is in the list, and False otherwise"""
    for element in list:
        if element == item:
            return True
    return False
```

Can make the code smaller, but still need to check each element

And sets check containment in *constant time* (i.e., time does not depend on list length)

How hashing makes sets magically fast

Definitions:

- *hash* (verb): To map a value to an integer index
- *hash table* (noun): A list that stores elements via hashing
- *hash set* (noun): A set of elements stored using the same hash function
- *hash function* (noun): An algorithm that maps values to indexes

One possible hash function for integers: `i % length`

index	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
value										

How hashing makes sets magically fast

One possible hash function for integers: `i % length`

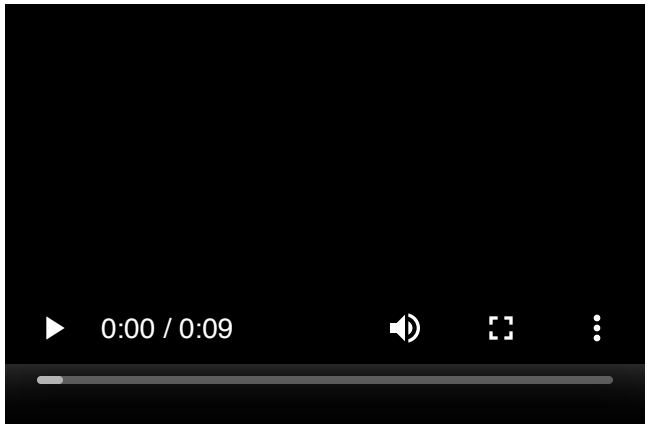
```
set.add(11)  # 11 % 10 == 1
set.add(49)  #  49 % 10 == 9
set.add(24)  #  24 % 10 == 4
set.add(7)  #  7 % 10 == 7
```

index	0	1	2	3	4	5	6	7	8	9
value		11			24			7		49

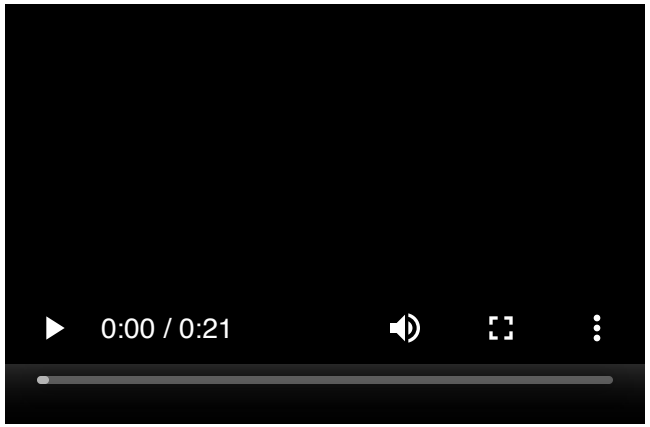
Okay, getting there, but there are some issues with that method...

dog hash collision gif

Source: https://m.vayagif.com/busqueda/0/el perro no nace agresivo/p/893



cat collision gif



no collision

Source: Tyler Yeats

How hashing makes sets magically fast

- *collision*: When hash function maps 2 values to same index

```
set.add(11)
set.add(49)
set.add(24)
set.add(7)

set.add(54)  # collides with 24
```

index	0	1	2	3	4	5	6	7	8	9
value		11			24			7		49

- *collision resolution* (noun): An algorithm for fixing collisions

How hashing makes sets magically fast

- *probing* (verb): Resolving a collision by moving to another index

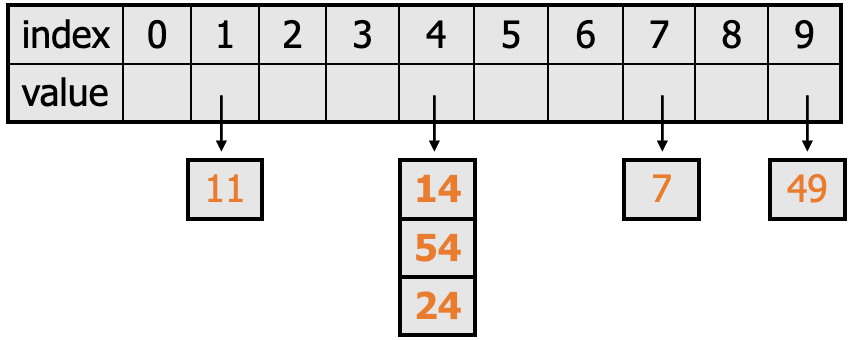
```
set.add(11)
set.add(49)
set.add(24)
set.add(7)

set.add(54)  # spot is taken -- probe (move to the next available spot)
```

index	0	1	2	3	4	5	6	7	8	9
value		11			24	54		7		49

How hashing makes sets magically fast

- *chaining* (verb): Resolving collisions by storing a list at each index
  - add / search / remove must traverse lists, but the lists are short
  - impossible to "run out" of indexes



Poll: Which of these can be done in constant time?

1. Checking if an element is in a set
2. Adding an element to a set
3. Removing an element from a set
4. Checking if a key is present in a map
5. Getting the value associated with a key in a map
6. Changing the value associated with a key in a map
7. Checking if a value appears in a map

What makes a good hash code?

`__hash__()` should return an `int` which is relatively unique to that object (so it can be used as an index in the hash table)

Rules for `__hash__()`:

- `__hash__()` must always return the same value for a given object
- If two objects are equal, then `__hash__()` must return the same value for them

Poll: Is this a legal hash function?

```
    def __hash__(self) -> int:
        return 42
```

1. Yes
2. No

In addition to the rules for hash functions...

Desired characteristics for `__hash__()`:

- We would like different objects to have different values
- The hash function should be quick to compute (ideally constant time)

Poll: Strings, numbers, and tuples are hashable by default in Python. Lists, sets, and dictionaries are not hashable by default in Python. Why might that be?

1. Because we rarely need to add lists, sets, or dictionaries to a set, or anything that requires hashing
2. Because lists, sets, and dictionaries are mutable, which could result in a changing hash code
3. Because lists, sets, and dictionaries are rarely equal to each other, so they don't need a hash code
4. Because lists, sets, and dictionaries can hold `None` in them, which shouldn't get a hash code

```python
from collections.abc import Hashable

class Course(Hashable):
    def __init__(self, department: str, course: int):
        self.department = department
        self.course = course

    def __str__(self) -> str:
        return f'{self.department}{self.course}'

    def __repr__(self) -> str:
        return self.__str__()

    def __eq__(self, other: object) -> bool:
        if not isinstance(other, Course):
            raise NotImplementedError
        else:
            return self.department == other.department and self.course == other.course

    def __hash__(self) -> int:
        return hash(str(self))


course_oakland = Course('CS', 2100)
course_boston = Course('CS', 2100)

courses: set[Course] = {course_oakland}
courses.add(course_boston)
```

Poll:

1. What is your main takeaway from today?

2. What would you like to revisit next time?