REPRESENTATION, MUTABLE TREES, LINKED LISTS Guide

CSM 61A

March 7 - March 11, 2022

Recommended Timeline

There are three big topics on this worksheet: representation, linked lists and mutable trees. There are plenty of problems, so poll your students about which topic to dive into; if no consensus, maybe present a little bit from each topic as time permits.

- Representation
- Representation Mini Lecture - 8 min
- Representation WWPD - 10 min (No need to cover all of it)
- Mutable Tree
- Mutable Trees Mini Lecture - 5 min
- Tree Sum - 5 min
- Double Tree - 10 min
- Linked Lists
- Linked Lists Lecture - 10 min
- WWPD - 10 min
- Skip - 5 min
- Skip (no mutate) - 10 min
- Has Cycle - 10 min

1 Representation

Representation Overview: `__repr__` and `__str__`

Classes can have ”magic methods” that add special built-in syntax features. They start and end with double underscores, such as in `__init__`. The goal of `__str__` is to convert an object to a human-readable string. The `__str__` function is helpful for printing objects and giving us information that's more readable than `__repr__`. Whenever we call **`print`**`()` on an object, it will call the `__str__` method of that object and print whatever value the `__str__` call returned. For example, if we had a `Person` class with a name instance variable, we can create a `__str__` method like this:

```
def __str__(self):
    return "Hello, my name is " + self.name
```

This `__str__` method gives us readable information: the person's name. Now, when we call print on a person, the following will happen:

```
>>> p = Person("John Denero")
>>> str(p)
'Hello, my name is John Denero'
>>> print(p)
Hello, my name is John Denero
```

The `__repr__` magic method returns the ”official” string representation of an object. You can invoke it directly by calling **`repr`**`(<some object>)`. However, `__repr__` doesn't always return something that is easily readable, that is what `__str__` is for. Rather, `__repr__` ensures that all information about the object is present in the representation. When you ask Python to represent an object in the Python interpreter, it will automatically call **`repr`** on that object and then print out the string that **`repr`** returns. If we were to continue our `Person` example from above, let's say that we added a **`repr`** method:

```
def __repr__(self):
    return "Name: " + self.name
```

Then we can write the following code:

```
# Python calls this object's repr function to see what
# to print on the line. Note, Python prints whatever
# result it gets from repr so it removes the quotes
# from the string
```

```
>>> p
Name: John Denero

# User is invoking the repr function directly.
# Since the function returns a string, its output
# has quotes. In the previous line, Python called
# repr and then printed the value. This line works
# like a regular function call: if a function
# returns a string, output that string with quotes.
>>> repr(p)
"Name: John Denero"
```

1. **Musician** - What would Python display? Write the result of executing the code and the prompts below. If a function is returned, write ”Function”. If nothing is returned, write ”Nothing”. If an error occurs, write ”Error”.

```
class Musician:
    popularity = 0
    def __init__(self, instrument):
        self.instrument = instrument
    def perform(self):
        print("a rousing " + self.instrument + " performance")
        self.popularity = self.popularity + 2
    def __repr__(self):
        return self.instrument

class BandLeader(Musician):
    def __init__(self):
        self.band = []
    def recruit(self, musician):
        self.band.append(musician)
    def perform(self, song):
        for m in self.band:
            m.perform()
        Musician.popularity += 1
        print(song)
    def __str__(self):
        return "Here's the band!"
    def __repr__(self):
        band = ""
        for m in self.band:
            band += str(m) + " "
        return band[:-1]

miles = Musician("trumpet")
goodman = Musician("clarinet")
ellington = BandLeader()
```

Some Quick Refreshers

Defining attributes: Instance attributes are defined with the `self.attr_name` notation (usually in `__init__` but could be elsewhere like in this problem). Class attributes are defined outside of methods in the body of the class definition, like the variable `popularity` in the class `Musician`.

Accessing attributes: Instance attributes are referred to using `self.attr_name` Class attributes can be referred to using `classname.attr_name` or `self.attr_name` (Note: using the latter will only work if there are no instance attributes bound with the name `attr_name`).

Before running any of the code below, `miles` and `goodman` are set to the musicians created as a result of calling the `__init__` constructor method in `Musician`. `ellington` uses `BandLeader's` `__init__` method, since `BandLeader` is the subclass and has `__init__` defined.

```
>>> ellington.recruit(goodman)
>>> ellington.perform()
```

Error

`ellington.recruit(goodman)` adds `goodman` to the end of `ellington's` instance attribute, `band`. Then, `ellington` checks its class (`BandLeader`) for the `perform()` method. But this `perform()` is expecting an argument, so this errors.

```
>>> ellington.perform("sing, sing, sing")
```

a rousing clarinet performance
sing, sing, sing

Using the same `perform()` method, now providing the correct number of arguments. First, going through the band list, `goodman` calls its `perform()` method, which is defined in `Musician`. Here, we print `"a rousing"` + `goodman`'s instrument + `" performance"`, and then `goodman`'s `self.popularity = self.popularity + 2` happens. The `self.popularity` on the right of the equal sign is `Musician.popularity` because `goodman` doesn't have its own instance attribute named `popularity` yet; then it becomes `self.popularity = 0 + 2`, and this creates the instance attribute `popularity` for `goodman`. Then `Musician.popularity`, the class attribute, in incremented by 1.

```
>>> goodman.popularity, miles.popularity
```

(2, 1)

First, we try to get the value of `goodman.popularity`. In our environment diagram, we see that `goodman` has the instance variable `popularity` already defined. Therefore, we get that value, 2, back. Then, we try to access `miles.popularity`. In this case, `miles` doesn't have a `popularity` instance variable defined, so we default to the class variable. There, we see it defined as 1, so we get that value. Finally, since commas in Python define a tuple, we return the two values as `(2, 1)`.

```
>>> ellington.recruit(miles)
>>> ellington.perform("caravan")
```

a rousing clarinet performance
a rousing trumpet performance
caravan

First, we call `ellington.recruit(miles)`. This appends `miles` to `ellington`'s instance variable, `band`. After that, we call `ellington.perform("caravan")`. Similar to the previous call on perform, we will loop through all of the values in `ellington.band`, calling their perform methods in order. This causes the first two lines to be printed. Next, we increment `Musician.popularity` (the class variable of `Musician` called `popularity`). Lastly, we print the `song` variable that was passed in, completing the last line.

```
>>> ellington.popularity, goodman.popularity, miles.popularity
```

(2, 4, 3)

```
>>> print(ellington)
```

Here's the band!

`print()` expects the string representation of `ellington`, which is given by calling the `__str__()` method of `ellington`. `ellington` checks to see if `BandLeader` has a `__str__()` method, which it does. So, `print(ellington)` then becomes `print("Here's the band!")`.

```
>>> ellington
```

clarinet trumpet

When prompting for `ellington`'s value, we return the representation of ellington given by `__repr__()`. So, we call `BandLeader`'s `__repr__()` method.

Teaching Tips

- For the error, it's important to make sure your students realize that BandLeader overrides Musician's `perform` function, and therefore a function call without the correct number of parameters in the new function will not work.
- Clarify to students the difference between `__str__` and `__repr__`, especially that **`print`** implicitly calls `__str__` and `__repr__`.
  - Another nuance to this which may confuse students is the `__repr__` method in BandLeader, which calls **`str`** on all Musicians in the band. Since the Musician class does not have a defined `__str__` method, it defaults to the defined `__repr__` method.
- The main challenge with this problem is distinctly identifying the class and instance variables and modifying both separately.
  - In particular, every Musician begins with a class variable `popularity`. However, after the first call to `perform`, a new instance variable `self.popularity` is created, which begins with the value `Musician.popularity` + 2.
  - After this first call to a Musician's `perform`, successive calls will increment the respective instance variable by 2.
  - Calling a BandLeader's `perform` function will increment the class variable `Musician.popularity`, which will raise the starting popularities of any new Musicians after their initial performances.
  - Ensure that students understand the interplay between the `popularity` class and instance variable.

2 Mutable Trees

For the following problems, use this definition for the Tree class:

```
class Tree:
    def __init__(self, label, branches=[]):
        self.label = label
        self.branches = list(branches)

    def is_leaf(self):
        return not self.branches

    def __repr__(self):
        if self.branches:
            branch_str = ', ' + repr(self.branches)
        else:
            branch_str = ''
        return 'Tree({0}{1})'.format(self.label, branch_str)
```

- The constructor constructs and returns a new instance of `Tree`

  `t = Tree(1)#creates a Tree instance with label 1 and no branches, since branches is defaulted to []`

- The `label` and `branches` are attributes, and `is_leaf()` is a method of the class.

  `t.label #returns the label of the tree`

  `t.branches #returns the branches of the tree, which is a list of trees`

  `t.is_leaf()#returns True if the tree is a leaf`

- A tree object is mutable

  To modify a `Tree` object, simply reassign its attributes. For example, `t.label = 2`.

  This means we can mutate values in the tree object instead of making a new tree that we return. In other words, we can solve tree class problems non-destructively and destructively.

1. Implement `tree_sum` which takes in a Tree object and replaces the label of the tree with the sum of all the values in the tree. `tree_sum` should also return the new label.

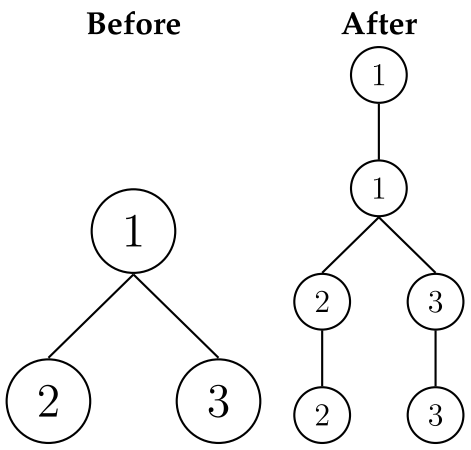
```
def tree_sum(t):
    """
    >>> t = Tree(1, [Tree(2, [Tree(3)]), Tree(4)])
    >>> tree_sum(t)
    10
    >>> t.label
    10
    >>> t.branches[0].label
    5
    >>> t.branches[1].label
    4
    """

    for b in t.branches:
        t.label += tree_sum(b)
    return t.label
```

Teaching Tips

- Make sure students understand why an explicit is_leaf() base case is unnecessary. If the function is called on a leaf, the for loop does not run, and it simply returns the label.
- The recursion occurs as part of the expression updating the label, which may confuse students at first. Explain how the returning of the label makes this work.
  - It may also help to show how the code would be written without tree_sum(b) on the right hand side of the expression to make the recursion clearer.
- Consider first drawing the Tree out and running through a doctest, showing how you would sum the labels in subtrees first before updating the root label.

2. DoubleTree hired you to architect one of their hotel expansions! As you might expect, their floor plan can be modeled as a tree and the expansion plan requires doubling each node (the patented double tree floor plan). Here's what some sample expansions look like:



Fill in the implementation for `double_tree`.

```
def double_tree(t):
    """
    Given a tree, return a new tree where entries appear
    twice.
    >>> double_tree(Tree(1))
    Tree(1, [Tree(1)])
    >>> double_tree(Tree(1, [Tree(2), Tree(3)]))
    Tree(1, [Tree(1, [Tree(2, [Tree(2)]),
                      Tree(3, [Tree(3)])
                     ])
            ])
    """

    if t.is_leaf():
        return Tree(t.label, [Tree(t.label)])
    else:
        dbl_branches = [double_tree(c) for c in t.branches]
        return Tree(t.label,
                    [Tree(t.label, dbl_branches)])
```

3 Linked Lists

Linked lists consists of a series of links which have two attributes: `first` and `rest`. The `first` attribute contains the value of the link (which can be an integer, string, list, even another linked list!). The `rest` attribute, on the other hand, is a pointer to another link or `Link.empty`, which is just an empty linked list represented traditionally by an empty tuple (but not necessarily, so never assume that it is represented by an empty tuple otherwise you will break an abstraction barrier!).

Because each link contains another link or `Link.empty`, linked lists lend themselves to recursion (just like trees). Consider the following example, in which we double every value in linked list. We mutate the current link and then recursively double the rest.

```
def double_values(link):
     if link is not Link.empty:
        link.first *= 2 # we mutate the value inside of the link
        double_val(link.rest) # we mutate the values in the rest
                              # of the linked list
    # if the link is empty then do nothing
```

However, unlike with trees, we can also solve many linked list questions using iteration. Take the following example where we have written `double_values` using a while loop instead of using recursion:

```
def double_values_iter(link):
    while link is not Link.empty:
        link.first *= 2
        link = link.rest # Note that this does not mutate
                         # the original linked list;
                         # it changes what link the variable
                         # link is pointing to
```

Teaching Tips

- Try to draw box and pointer diagrams.
- Make clear that the pointer *points* to a linked list if we have nested linked lists.
- Try to experiment with going over various ways to mutate and create linked lists.
- We have a great visualizer on `https://code.cs61a.org/` where you can call draw(lst) to visualize a list!
- Try using PythonTutor as well!

For each of the following problems, assume linked lists are defined as follows:

```
class Link:
    empty = ()
    def __init__(self, first, rest=empty):
        assert rest is Link.empty or isinstance(rest, Link)
        self.first = first
        self.rest = rest

    def __repr__(self):
        if self.rest is not Link.empty:
            rest_repr = ', ' + repr(self.rest)
        else:
            rest_repr = ''
        return 'Link(' + repr(self.first) + rest_repr + ')'

    def __str__(self):
        string = '<'
        while self.rest is not Link.empty:
            string += str(self.first) + ' '
            self = self.rest
        return string + str(self.first) + '>'
```

To check if a `Link` is empty, compare it against the class attribute `Link.empty`:

```
if link is Link.empty:
    print('This linked list is empty!')
```

1. What will Python output? Draw box-and-pointer diagrams to help determine this.

```
>>> a = Link(1, Link(2, Link(3)))

+---+---+  +---+---+  +---+---+
| 1 | --|->| 2 | --|->| 3 | / |
+---+---+  +---+---+  +---+---+

>>> a.first

1

>>> a.first = 5

+---+---+  +---+---+  +---+---+
| 5 | --|->| 2 | --|->| 3 | / |
+---+---+  +---+---+  +---+---+

>>> a.first
```

5

```
>>> a.rest.first
```

2

```
>>> a.rest.rest.rest.rest.first
```

Error: tuple object has no attribute rest (Link.empty has no rest)

```
>>> a.rest.rest.rest = a

   +---+---+   +---+---+   +---+---+
+->| 5 | --|->| 2 | --|->| 3 | --|--+
|  +---+---+   +---+---+   +---+---+  |
|                                     |
+-------------------------------------+

>>> a.rest.rest.rest.rest.first

2

>>> repr(Link(1, Link(2, Link(3, Link.empty))))

"Link(1, Link(2, Link(3)))"

>>> Link(1, Link(2, Link(3, Link.empty)))

Link(1, Link(2, Link(3)))

>>> str(Link(1, Link(2, Link(3))))

'<1 2 3>'

>>> print(Link(Link(1), Link(2, Link(3))))

<<1> 2 3>
```

Teaching Tips

- For assignment statements, Python will not print anything but still have them draw out what the linked list will look like
- Note that we are doing mutation here, so we are actually altering the object that was created in the first assignment.
  - Some students may have minimal exposure to mutating objects so try to emphasize this and make it obvious through diagrams.
- For the error, walk-through how to keep track of which rest corresponds to which object in the box and pointer diagram. **Make sure they understand why calling rest a fourth time will give us an error (look back at the class definition)**
  - Abstraction:
    * our last .rest is set to Link.empty
    * Link.empty is not a Link objects — they do not have a .rest attribute
  - Actual implementation:
    * our last .rest is set to Link.empty

    * Link.empty is not a Link objects — they do not have a .rest attribute

- Reassigning the last .rest to point back at the front always trips students up.
  - Make it clear that a is a pointer that points to the linked list. So we are trying to assign the last rest of a to point at what a points to, which is the beginning of the list. **To test their understanding ask what would be different if we instead had**:
    * a.rest.rest.rest = a.rest
  - a way to explain the assignment for this problem is to emphasize the "evaluation" of the RHS and the LHS
  - what is the value of a (a pointer). Really emphasize the implications of pointers here.
  - where are we putting a into? (the box that represents a.rest.rest.rest)
  - same for a.rest.rest.rest = a.rest. what is the value of a.rest? (still a pointer!)
  - Mention that this creates a cycle in the list

2. Write a function `skip`, which takes in a `Link` and returns a new `Link` with every other element skipped.

```
def skip(lst):
    """
    >>> a = Link(1, Link(2, Link(3, Link(4))))
    >>> a
    Link(1, Link(2, Link(3, Link(4))))
    >>> b = skip(a)
    >>> b
    Link(1, Link(3))
    >>> a
    Link(1, Link(2, Link(3, Link(4)))) # Original is unchanged
    """
    if ___________________________________________:

        ___________________________________________

    elif _________________________________________:

        ___________________________________________

    _______________________________________________
```

```
    if lst is Link.empty
        return Link.empty
     elif lst.rest is Link.empty:
        return Link(lst.first)
    return Link(lst.first, skip(lst.rest.rest))
```

Base cases:

- When the linked list is empty, we want to return a new Link.empty.
- If there is only one element in the linked list (aka the next element is empty), we want to return a new linked list with that single element.

Recursive case:
All other longer linked lists can be reduced down to either a single element or empty linked list depending on whether it has odd or even length. Therefore, we want to keep the first element, and recurse on the element after the next (skipping the immediate next element with `lst.rest.rest`). To build a new linked list, we can add new links to the end of the linked list by calling skip recursively inside the `rest` argument of the `Link` constructor.

Teaching Tips

- Walk through what we want to do by looking at an example box-and-pointer diagram first.
- Make sure they understand, in English, what we are trying to do.
- If students are struggling, have them think about what we can change (pointers), since we can't make new Link objects
  - Specifically, compare the pointers in the original list to the ones in the output list.
  - Think about how you could modify the original pointers.

3. Now write function `skip` by mutating the original list, instead of returning a new list. Do NOT call the `Link` constructor.

```
def skip(lst):
    """
    >>> a = Link(1, Link(2, Link(3, Link(4))))
    >>> skip(a)
    >>> a
    Link(1, Link(3))
    """
```

```
def skip(lst): # Recursively
    if lst is Link.empty or lst.rest is Link.empty:
        return
    lst.rest = lst.rest.rest
    skip(lst.rest)
```

```
def skip(lst): # Iteratively
    while lst is not Link.empty and lst.rest is not Link.empty
      :
        lst.rest = lst.rest.rest
        lst = lst.rest
```

Because this problem is mutative, we should never be creating a new list - we should never have `Link(x)`, or the creation of a new Link instance, anywhere in our code! Instead, we'll be reassigning `lst.rest`.

In order to skip a node, we can assign `lst.rest = lst.rest.rest`. If we have lst assigned to a link list that looks like the following:

```
1 -> 2 -> 3 -> 4 -> 5
```

Setting `lst.rest = lst.rest.rest` will take the arrow that points form 1 to 2 and change it to point from 1 to 3. We can see this by evaluating `lst.rest.rest`. `lst.rest` is the arrow that comes from 1, and `lst.rest.rest` is the link with 3.

Once we've created the following list:

```
1 -> 3 -> 4 -> 5
```

we just need to call skip on the rest of the list. If we call skip on the list that starts at 3, we'll skip over the link with 4 and set the pointer from 3 to point to the link with 5. This is the behavior that we want! Therefore, our recursive call is `skip(lst.rest)`, since `lst.rest` is now the link that contains 3.

Teaching Tips

- Make sure they understand when we are mutating and when we are creating a new linked list
- Draw box-and-pointer diagrams!
- Look for "patterns" or repeated work while you work with your box-and-pointer diagram that you can abstract away with your recursive call.
- Sometimes it is easier to write the recursive call before doing the base cases
- I usually write the recursive call and then see what could "break"
  - If we access lst.first at any point, we have to make sure that lst exists
  - If we access lst.rest.rest at any point we have to make sure that lst.rest exists
  - What errors would we get if we didn't ensure these conditions?

4. **(Optional)** Write `has_cycle` which takes in a `Link` and returns `True` if and only if there is a cycle in the `Link`. Note that the cycle may start at any node and be of any length. Try writing a solution that keeps track of all the links we've seen. Then try to write a solution that doesn't store those witnessed links (consider using two pointers!).

```python
def has_cycle(s):
    """
    >>> has_cycle(Link.empty)
    False
    >>> a = Link(1, Link(2, Link(3)))
    >>> has_cycle(a)
    False
    >>> a.rest.rest.rest = a
    >>> has_cycle(a)
    True
    """

    if s is Link.empty:
        return False
    slow, fast = s, s.rest
    while fast is not Link.empty:
        if fast.rest is Link.empty:
            return False
        elif fast is slow or fast.rest is slow:
            return True
        slow, fast = slow.rest, fast.rest.rest
    return False
```

Teaching Tips

- Go through multiple examples of Linked List with cycles alongside examples of Linked Lists without cycles.
- Ask your students what patterns they see for lists that have cycles
- It might take some time for students to come up with the fast and slow pointers solution. A common analogy used is the hare and tortoise analogy for this problem.
- If the slow pointer catches up to the fast pointer, we know a cycle must have occured because the slow pointer should never pass the fast pointer in a non-cycle list.