Dictionaries

Dictionaries are like sets, except that the "elements" of the dictionary are treated as keys, and a value is associated with that key. As in sets, the keys to dictionaries must be hashable objects. However, the values associated with the key can be anything, and can be mutable. In addition, the association between key and value can also be changed.

In [1]:
```
table = {} #Create empty dictionary. dict() also works
table[27] = 'my value'
table["dog"] = [1, 2, "three"]
print(table)
```

```
{27: 'my value', 'dog': [1, 2, 'three']}
```

In [2]:
```
table[27] = 3
print("table:", table)
```

```
table: {27: 3, 'dog': [1, 2, 'three']}
```

In [3]:
```
#Is key in a dictionary? Can use 'in' syntax.
if 'dog' in table:
    table['cat'] = 'unhappy'
print("table:", table)
```

```
table: {27: 3, 'dog': [1, 2, 'three'], 'cat': 'unhappy'}
```

In [4]:
```
# Iterate over keys in a dictionary. Any order can occur!
for key in table:
    print("key:", key, "val:", table[key])
```

```
key: 27 val: 3
key: dog val: [1, 2, 'three']
key: cat val: unhappy
```

In [5]:
```
# Also, can iterate over key, val items in a dictionary. Any order can occur!
for key, val in table.items():
    print("key:", key, "; val:", val)
```

```
key: 27 ; val: 3
key: dog ; val: [1, 2, 'three']
key: cat ; val: unhappy
```

In [6]:
```
# Remove element from a dictionary
del table[27]  #Exception if key is not in dictionary
print("table:", table)
```

```
table: {'dog': [1, 2, 'three'], 'cat': 'unhappy'}
```

In [7]:
```
# Dictionary comprehensions also possible
{n: n**3 for n in range(8)}
```

Out[7]:
```
{0: 0, 1: 1, 2: 8, 3: 27, 4: 64, 5: 125, 6: 216, 7: 343}
```

Some fun with sets and dictionaries...

Problem:

Return a new list with the 2nd instance of the first element that is repeated in the input list removed. The rest of the list should remain unchanged (be the same as the input).

Goal:

Our goal is to use this problem to build some familiarity with **sets and dictionaries** -- so the code versions below are written to use sets and/or dictionaries. This is definitely not the only way to solve the problem, but it happens to also be the most efficient away. (At the very end, we'll also show a more "direct", but less efficient, solution which does not use sets or dictionaries.)

In [8]:
```
inp = [0, 12, 12, 0, 12, 12, 34, 56, 23, 11, 45, 2, 3, 4, 7, 10, 12]

inp_str = ['zero', 'twelve', 'twelve', 'zero', 'twelve', 'twelve',
           'thirty four', 'fifty six', 'twenty three', 'eleven',
           'forty five', 'two', 'three', 'four', 'seven', 'ten', 'twelve']
```

Version 1

In [9]:
```
def remove_2nd_instance_v1(data):
    print("data:", data)

    #Create a dictionary with frequencies
    freqd = {}
    for x in data:
        freqd[x] = freqd.get(x, 0) + 1
        #The get() method returns the value of the key in the dictionary
        #if it exists, and otherwise returns the default value (2nd param)

    #Look for first element in list, that is repeated somewhere later
    repeated = None
    for x in data:
        if freqd[x] >= 2:
            repeated = x
            break
    print("first repeated:", repeated)

    #Look through the list to find where the 2nd instance of the repeat is
    index = len(data)
    count = 0
    for i in range(len(data)):
        if data[i] == repeated:
            count += 1
        if count == 2:
            index = i
            break
    print("index:", index)

    output = data[:]        #make copy of data
    #Remove this 2nd instance if it exists from a copy of the input
    if index < len(output):
        output.pop(index)
    return output

remove_2nd_instance_v1(inp)
```

```
data: [0, 12, 12, 0, 12, 12, 34, 56, 23, 11, 45, 2, 3, 4, 7, 10, 12]
first repeated: 0
index: 3
```

Out[9]:
```
[0, 12, 12, 12, 12, 34, 56, 23, 11, 45, 2, 3, 4, 7, 10, 12]
```

In [10]:
```
output = remove_2nd_instance_v1(inp_str)  #works also for lists of other kinds of elements
print(output)
```

```
data: ['zero', 'twelve', 'twelve', 'zero', 'twelve', 'twelve', 'thirty four', 'fifty six', 'twen
ty three', 'eleven', 'forty five', 'two', 'three', 'four', 'seven', 'ten', 'twelve']
first repeated: zero
index: 3
['zero', 'twelve', 'twelve', 'twelve', 'twelve', 'thirty four', 'fifty six', 'twenty three', 'el
even', 'forty five', 'two', 'three', 'four', 'seven', 'ten', 'twelve']
```

Revised Spec

Our procedure removed `0` (`'zero'`) from the list at index 3, since that is the first repeat of the "earliest" element that has a later repeat. Is that what we want? Our specification is perhaps **ambiguous**. Arguably, we might want to remove `12` at index 2 since `12` appeared twice before `0` appeared twice. Let's clarify our spec: by 'first repeated element' we mean the element that appears twice **first**. Thus
`inp = [0, 12, 12, 0, 12, 12, 34, 56, 23, 11, 45, 2, 3, 4, 11, 10, 12]`

should give
`expect = [0, 12, 0, 12, 12, 34, 56, 23, 11, 45, 2, 3, 4, 11, 10, 12] # remove second 12`

Version 2 -- remove the second appearance of the first element that makes a second appearance

In [11]:
```
def remove_2nd_instance_v2(data):
    print("data:", data)

    # Creat_e a dictionary whose values are frequencies AND indices of elements,
    # i.e., [freq_of_val, index_of_first_appearance_or_of_second_appearance_if_repeated].
    # We store either the index of the first element, or for repeated elements
    # we store the index of the 2nd instance (first repeat). Ignore additional repeats.
    freqd = {}

    for i in range(len(data)):
        x = data[i]
        if x in freqd:
            freqd[x][0] += 1
            if freqd[x][0] == 2:
                freqd[x][1] = i
        else:
            freqd[x] = [1, i]
    print("freqd =", freqd)

    #Get the index of the 2nd instance of the earliest appearing repeated element.
    index = len(data)
    for x in data:
        entry = freqd[x]
        if entry[0] >= 2:
            index = min(index, entry[1])
    print("index:", index)

    #Remove this 2nd instance
    output = data[:]        #make copy of data
    if index < len(output):
        output.pop(index) #list pop -- mutates list
    return output

remove_2nd_instance_v2(inp)
```

```
data: [0, 12, 12, 0, 12, 12, 34, 56, 23, 11, 45, 2, 3, 4, 7, 10, 12]
freqd = {0: [2, 3], 12: [5, 2], 34: [1, 6], 56: [1, 7], 23: [1, 8], 11: [1, 9], 45: [1, 10], 2:
[1, 11], 3: [1, 12], 4: [1, 13], 7: [1, 14], 10: [1, 15]}
index: 2
```

Out[11]:
```
[0, 12, 0, 12, 12, 34, 56, 23, 11, 45, 2, 3, 4, 7, 10, 12]
```

Hurray! We correctly detect the second `12` as the first repeated element, and remove that.

But can we do better in terms of writing prettier and tighter code -- perhaps something using **sets** rather than **dictionaries**?

Version 3 -- simplified using a set rather than a dictionary, to detect first repeat

In [12]:
```
def remove_2nd_instance_v3(data):
    print("data:", data)

    index = len(data)
    repeated = set()
    for index, x in enumerate(data):
        if x in repeated:
            break
        repeated.add(x)
    print("freq_set:", repeated)
    print("index:", index)

    #Remove this 2nd instance
    output = data[:index] + data[index+1:]  # list slicing rather than list pop
    return output

remove_2nd_instance_v3(inp)
```

```
data: [0, 12, 12, 0, 12, 12, 34, 56, 23, 11, 45, 2, 3, 4, 7, 10, 12]
freq_set: {0, 12}
index: 2
```

Out[12]:
```
[0, 12, 0, 12, 12, 34, 56, 23, 11, 45, 2, 3, 4, 7, 10, 12]
```

Version 4 -- more direct but less efficient

Here's a very short solution to the original problem (with the first interpretation of the spec). It's interesting to think about how this works, and its virtues and/or failings compared to the earlier versions. This operates directly on the input list, without making any auxiliary sets and dictionaries, so in some sense is a more direct solution of the earlier problem. But it doesn't use sets and dictionaries, and as a result, it is slow for large inputs (taking quadratic time instead of linear). Think about why!

In [13]:
```
def remove_2nd_instance_v4(data):
    print("data:", data)
    for i in range(len(data)):
        try:
            ii = data.index(data[i],i+1) #raises exception if data[i] not found later in list
            print("ii:", ii)
            return data[:ii] + data[ii+1:]
        except ValueError:
            continue
    return data[:]

remove_2nd_instance_v4(inp)
```

```
data: [0, 12, 12, 0, 12, 12, 34, 56, 23, 11, 45, 2, 3, 4, 7, 10, 12]
ii: 3
```

Out[13]:
```
[0, 12, 12, 12, 12, 34, 56, 23, 11, 45, 2, 3, 4, 7, 10, 12]
```

Consider it a challenge to you to revise this to go with our revised spec, i.e., to remove the first repeated element (12)!