Phát triển phần mềm nâng cao cho tính toán khoa học Cấu trúc dữ liêu cơ bản

Vũ Tiến Dũng

Khoa Toán - Cơ - Tin học Trường ĐH Khoa học Tự Nhiên Hà Nội

Nội dung

- 1 Xâu ký tự String
- 2 Danh sách List
- 3 Bộ Tuple
- 4 Tập hợp Set
- 5 Từ điển Dictionary

- Kiểu dữ liệu xâu ký tự (string) trong Python là một mảng của các byte biểu diễn các ký tự unicode.
- Tuy nhiên, Python không có kiểu dữ liệu riêng để biểu diễn ký tự, một ký tự trong xâu đơn giản là một xâu độ dài một.
- Một xâu là một mảng các ký tự, các phần tử trong mảng có thể được tham chiếu qua chỉ số đặt trong ngoặc vuông [].

```
>>> a = 'Hello world'
>>> print(a)
Hello world
>>> type(a)
<class 'str'>
>>> print(a[1])
e
>>> type(a[1])
<class 'str'>
>>>
```

• Các phần tử trong xâu có thể được truy cập theo lát (Slice).

```
b = "Hello, World!"
print(b[2:5])
```

• Chú ý chỉ số các phần tử trong xâu, hay trong mảng nói chung được tính từ 0, trong ví dụ trên phần tử có chỉ số 2 là I, phần tử có chỉ số 5 là dấu, kết quả là IIo. Chú ý kết quả chỉ lấy đến trước phần tử có chỉ số 5.

Hình 1: Tham chiếu các phần tử trong mảng (danh sách - list) theo lát (slice)

```
>>> b = "Hello, World"
>>> b[2:5]
'11o'
>>> b[2:-1]
'llo, Worl'
>>> b[-1:2]
. .
>>> b[2:-1:2]
'lo ol'
>>> b[::2]
'Hlo ol'
>>> b[:5]
'Hello'
>>> b[5:]
 , World'
```

Lấy độ dài xâu len()

```
a = "Hello, World!"
print(len(a))
```

Xóa khoảng trắng thừa strip()

```
a = " Hello, World! "
print(a.strip()) # returns "Hello, World!"
```

Chuyển ký tự về ký tự hoa / thường lower() / upper()

```
a = "Hello, World!"

print(a.lower())

print(a.upper())
```

Thay thế xâu replace()

```
a = "Hello, World!"
print(a.replace("H", "J"))
```

Tách xâu split()

```
a = "Hello, World!"
print(a.split(",")) # returns ['Hello', ' World!']
```

Kiểm tra xâu có chứa / Không chứa xâu con in / not in

```
txt = "The rain in Spain stays mainly in the plain"

x = "ain" in txt

print(x)

txt = "The rain in Spain stays mainly in the plain"

x = "ain" not in txt

print(x)
```

Cộng xâu / nối xâu

```
1     a = "Hello"
2     b = "World"
3     c = a + b
4     print(c)
```

 Chú ý việc cộng xâu chỉ thực hiện giữa 2 xâu ký tự, nếu muốn thực hiện cộng xâu với số, ta phải chuyển số sang dạng xâu ký tự, tương tự với các kiểu dữ liệu khác

```
year = 2019
print("Happy new year "+str(year))
```

Định dạng xâu format()

```
age = 36
        txt = "My name is John, and I am {}"
        print(txt.format(age))
3
        quantity = 3
4
5
        itemno = 567
        price = 49.95
6
        myorder = "I want {} pieces of item {} for {}
       dollars."
8
        print(myorder.format(quantity, itemno, price))
9
        quantity = 3
10
        itemno = 567
11
        price = 49.95
12
        myorder = "I want to pay {2} dollars for {0} pieces
13
       of item {1}."
        print(myorder.format(quantity, itemno, price))
14
```

• Các phương thức thường dung cho xâu ký tự - string

```
capitalize(); casefold(); center(); count();
encode(); endswith(); expandtabs(); find(); format();
format_map(); index(); isalnum(); isalpha();
isdecimal(); isdigit(); isidentifier(); islower();
isnumeric(); isprintable(); isspace(); istitle();
isupper(); join(); ljust(); lower(); lstrip();
maketrans(); partition(); replace(); rfind(); rindex();
rjust(); rpartition(); rsplit(); rstrip(); split();
splitlines(); startswith(); strip(); swapcase();
title(); translate(); upper(); zfill();
```

 Một danh sách là một bộ có tính thứ tự và có thể thay đổi được. Trong Python một danh sách được đặt trong một cặp ngoặc vuông []

```
thislist = ["apple", "banana", "cherry"]
print(thislist)
```

 Các phần tử trong danh sách được truy cập đến thông qua chỉ số của phần tử đó. Chỉ số được đánh bắt đầu từ 0, và xoay vòng (phần tử cuối cùng tương ứng với chỉ số -1)

```
thislist = ["apple", "banana", "cherry"]
print(thislist[1])
print(thislist[-1])
```

Truy cập đến một vùng các phần tử theo chỉ số

```
thislist = ["apple", "banana", "cherry", "orange",

"kiwi", "melon", "mango"]

print(thislist[2:5])#['cherry', 'orange', 'kiwi']

print(thislist[-4:-1]) #['orange', 'kiwi', 'melon']
```

Thay đổi giá trị của phần tử trong danh sách

```
thislist = ["apple", "banana", "cherry"]
thislist[1] = "blackcurrant"
print(thislist)
```

Duyệt qua các phần tử trong danh sách với vòng lặp for

```
thislist = ["apple", "banana", "cherry"]
for x in thislist:
    print(x)
```

Phần tử được thêm vào cuối danh sách bằng lệnh append()

```
thislist = ["apple", "banana", "cherry"]
thislist.append("orange")
print(thislist)
```

• Để thêm phần tử vào vị trí cụ thế dùng lệnh insert()

```
thislist = ["apple", "banana", "cherry"]
thislist.insert(1, "orange")
print(thislist)
```

Xóa phần tử khỏi danh sách remove()

```
thislist = ["apple", "banana", "cherry"]
thislist.remove("banana")
print(thislist)
```

- Lấy phần tử và xóa khỏi danh sách pop()
- Mặc định thì lệnh pop() xóa phần tử ở cuối danh sách nếu không có tham số về chỉ số phần tử muốn xóa.

```
thislist = ["apple", "banana", "cherry"]
thislist.pop()
print(thislist)
```

• lệnh pop() cũng được sử dụng để lấy và xóa phần tử ở một vị trí cụ thể.

```
thislist = ["apple", "banana", "cherry"]
thislist.pop(1)
print(thislist)
```

Xóa phần tử trong danh sách del

```
thislist = ["apple", "banana", "cherry"]

del thislist[0]

print(thislist)
```

Làm trống danh sách clear()

```
thislist = ["apple", "banana", "cherry"]
thislist.clear()
print(thislist)
```

Sao chép danh sách copy()

```
thislist = ["apple", "banana", "cherry"]
mylist = thislist.copy()
print(mylist)
```

Hình 2: Ví dụ về copy danh sách

```
>>> thislist = ["apple", "banana", "cherry"]
>>> mylist = thislist.copy()
>>> print(mylist)
['apple', 'banana', 'cherry']
>>> mylist.append(3)
>>> print(mylist)
['apple', 'banana', 'cherry', 3]
>>> print(thislist)
['apple', 'banana', 'cherry']
```

Chú ý phép gán

```
list2 = list1
```

 Chỉ là phép gán tham chiếu, kho đó việc thay đổi 1 trong 2 danh sách thì danh sách kia cũng thay đổi theo.

```
>>> list = [1,2,3,4]
>>> list1 = [1,2,3,4]
>>> list2 = list1
>>> list2.append(5)
>>> print(list1)
[1, 2, 3, 4, 5]
>>> print(list2)
[1, 2, 3, 4, 5]
>>> list1.append(5)
>>> print(list2)
[1, 2, 3, 4, 5, 5]
>>> print(list2)
```

Kết hợp hai danh sách

```
list1 = ["a", "b" , "c"]
list2 = [1, 2, 3]
list3 = list1 + list2
print(list3)
```

 Một cách khác có thể để kết hợp 2 danh sách là dùng lệnh extend()

```
list1 = ["a", "b" , "c"]
list2 = [1, 2, 3]
list1.extend(list2)
print(list1)
```

Hàm dựng của danh sách list()

```
thislist = list(("apple", "banana", "cherry")) # note
the double round-brackets
print(thislist)
```

Danh sách - List, Ví dụ

```
def sumEven(a):
        return sum(a[i] for i in range(0, len(a)) if a[i]%2 == 0)
        def sumOdd(a):
4
        return sum(i for i in a if i%2 == 1)
5
6
        def sumList(a):
7
        return sum(a)
8
9
        n = int(input('Enter the number of elements in array n =
10
       '))
        aList = []
        for i in range(0, n):
12
        aList.append( int(input('list[' + str(i) + '] = ')))
13
        print('Sum of even numbers in array = ', sumEven(aList))
14
        print('Sum of odd numbers in array = ', sumOdd(aList))
15
        print('Sum of all elements in array = ', sumList(aList))
16
```

Danh sách - List, Ví dụ

```
aList = [0 for j in range(0, n)]

for i in range(0, n):
    aList[i] = int(input('list[' + str(i) + '] = '))
```

Danh sách - List, Ví dụ

```
def wordCount(s):
      tokens = s.split(' ')
      words = list(dict.fromkeys(tokens))
      counts = [tokens.count(words[i]) for i in range(0,
4
      len(words))]
      return words, counts
5
6
  text = "the quick brown fox jumps over the lazy dog "
8
  words, counts = wordCount(text)
10
  for i in range(0,len(words)):
      print(words[i],':',counts[i])
12
```

```
1 >>> fruits = ['orange', 'apple', 'pear', 'banana', 'kiwi',
      'apple', 'banana']
2 >>> fruits.count('apple')
4 >>> fruits.count('tangerine')
  0
6 >>> fruits.index('banana')
  3
8 >>> fruits.index('banana', 4) # Find next banana starting a
      position 4
  6
9
10 >>> fruits.append('grape')
11 >>> fruits
12 ['banana', 'apple', 'kiwi', 'banana', 'pear', 'apple', 'orange',
      'grape']
13 >>> fruits.sort()
14 >>> fruits
['apple', 'apple', 'banana', 'banana', 'grape', 'kiwi',
      'orange', 'pear']
16 >>> fruits.pop() # 'pear'
```

Sử dụng danh sách như ngăn xếp

```
| >>>  stack = [3, 4, 5]
2 >>> stack.append(6)
3 >>> stack.append(7)
4 >>> stack
5 [3, 4, 5, 6, 7]
6 >>> stack.pop()
7 7
8 >>> stack
9 [3, 4, 5, 6]
10 >>> stack.pop()
  6
11
12 >>> stack.pop()
13 5
14 >>> stack
15 [3, 4]
```

Sử dụng danh sách làm hàng đợi

```
>>> from collections import deque
>>> queue = deque(["Eric", "John", "Michael"])
>>> queue.append("Terry") # Terry arrives
>>> queue.append("Graham") # Graham arrives
>>> queue.popleft() # The first to arrive now leaves
'Eric'
>>> queue.popleft() # The second to arrive now leaves
'John'
>>> queue # Remaining queue in order of arrival
deque(['Michael', 'Terry', 'Graham'])
```

 Bộ là một tập hợp có thứ tự và không thể thay đổi, trong Python bộ được khai báo trong cặp ngoặc đơn ().

```
thistuple = ("apple", "banana", "cherry")
print(thistuple) #>>>('apple', 'banana', 'cherry')
```

 Truy cập đến các phần tử của bộ cũng giống như việc truy cập đến các phần tử trong danh sách.

Xâu ký tự - String

 Khi một bộ được tạo ra thì các giá trị của các phần tử trong bộ không thể thay đổi được, tuy nhiên chúng ta có thể chuyển đổi bộ sang danh sách để thực hiện các thay đổi nếu cần thiết.

```
x = ("apple", "banana", "cherry")
y = list(x)
y[1] = "kiwi"
x = tuple(y)
print(x) #>>>('apple', 'kiwi', 'cherry')
```

Duyệt qua các phần tử trong bộ

```
thistuple = ("apple", "banana", "cherry")
for x in thistuple:
   print(x)
```

Kiểm tra phần tử có trong bộ hay không

```
thistuple = ("apple", "banana", "cherry")
if "apple" in thistuple:
print("Yes, 'apple' is in the fruits tuple")
```

Kích thước của bộ

```
thistuple = ("apple", "banana", "cherry")
print(len(thistuple))
```

Tạo bộ có một phần tử

```
thistuple = ("apple",)
print(type(thistuple)) #>>><class 'tuple'>

#NOT a tuple
thistuple = ("apple")
print(type(thistuple)) #>>><class 'str'>
```

Xóa bộ

```
thistuple = ("apple", "banana", "cherry")
del thistuple
print(thistuple) #this will raise an error
```

Gôp bộ

```
tuple1 = ("a", "b", "c")
tuple2 = (1, 2, 3)

tuple3 = tuple1 + tuple2
print(tuple3) #>>>('a', 'b', 'c', 1, 2, 3)
```

Hàm dựng

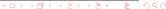
```
thistuple = tuple(("apple", "banana", "cherry"))
# note the double round-brackets
print(thistuple)
```

Đếm số phần tử có trong bộ

```
print(thistuple.count("cherry"))
```

• Tìm vị trí của phần tử trong bộ

```
print(thistuple.index("cherry"))
```



Một số ví dụ về bộ - tuple

```
1 >>>t = 12345, 54321, 'hello!'
2 >>>type(t)
3 <class 'tuple'>
4 >>> t = (12345, 54321, 'hello!')
5 >>> # Tuples may be nested:
[6] ... u = t, (1, 2, 3, 4, 5)
7 >>> u
8 ((12345, 54321, 'hello!'), (1, 2, 3, 4, 5))
9
10 >>> # Tuples are immutable:
11 ... t[0] = 88888
12 Traceback (most recent call last):
File "<stdin>", line 1, in <module>
14 TypeError: 'tuple' object does not support item assignment
15 >>> # but they can contain mutable objects:
|16| \dots v = ([1, 2, 3], [3, 2, 1])
17 >>> V
18 ([1, 2, 3], [3, 2, 1])
```

Một số ví dụ về bộ - tuple

```
1 >>> empty = ()
2 >>> singleton = 'hello', # <-- note trailing comma
3 >>> len(empty)
4 0
5 >>> len(singleton)
6 1
7 >>> singleton
8 ('hello',)
```

Tập hợp là một bộ không có thứ tự và không có chỉ số, trong
 Python một tập hợp được đặt trong cặp ngoặc nhọn {}.

```
thisset = {"apple", "banana", "cherry"}
print(thisset)
for x in thisset:
    print(x)
print("banana" in thisset)
```

- Khi được tạo ra các phần tử trong tập hợp không thể thay đổi giá trị, nhưng có thể thêm phần tử mới vào tập hợp
- Để thêm một phần tử vào tập hợp sử dụng phương thức add()

```
thisset = {"apple", "banana", "cherry"}
thisset.add("orange")
print(thisset)
```

 Để thêm nhiều phần tử vào tập hợp ta sử dụng phương thức update()

```
thisset = {"apple", "banana", "cherry"}
thisset.update(["orange", "mango", "grapes"])
print(thisset)
```

Kích thước tập hợp

```
thisset = {"apple", "banana", "cherry"}
print(len(thisset))
```

Xóa phần tử khỏi tập hợp remove()/discard()

```
thisset = {"apple", "banana", "cherry"}
thisset.remove("banana")
print(thisset)
thisset.discard("cherry")
print(thisset)
```

Xâu ký tự - String

• Phương thức pop() cũng có thể được sử dụng để xóa phần tử khỏi tập hợp, tuy nhiên phương thức pop() sẽ xóa phần tử cuối cùng nhưng do tập hợp không có tính thứ tự nên chúng ta không biết được phần tử nào sẽ bị xóa khỏi tập hợp

```
thisset = {"apple", "banana", "cherry"}
x = thisset.pop()
print(x)
print(thisset)
```

Phương thức clear() xóa tất cả các phần tử trong tập hợp

```
thisset = {"apple", "banana", "cherry"}
thisset.clear()
print(thisset)

# The del method deletes the entire set from memory
thisset = {"apple", "banana", "cherry"}
del thisset
print(thisset)
```

Hợp hai tập hợp union()/update()

```
set1 = {"a", "b" , "c"}
set2 = {1, 2, 3}
set3 = set1.union(set2)
print(set3)
set1.update(set2)
print(set1)
```

Hàm dựng của tập hợp

```
thisset = set(("apple", "banana", "cherry"))
```

Một số ví dụ trên Tập hợp - Set

```
1 >>> a = {x for x in 'abracadabra' if x not in 'abc'}
2 >>> a
3 {'r', 'd'}
5 >>> basket ={'apple', 'orange', 'apple', 'pear', 'orange',
       'banana'}
6 >>> print(basket) # show that duplicates have been removed
7 {'orange', 'banana', 'pear', 'apple'}
8 >>> 'orange' in basket # fast membership testing
9 True
10 >>> 'crabgrass' in basket
11 False
```

Một số ví dụ trên Tập hợp - Set

```
1 >>> a = set('abracadabra')
2 >>> b = set('alacazam')
3 >>> a # unique letters in a
4 {'a', 'r', 'b', 'c', 'd'}
5 >>> a - b # letters in a but not in b
6 {'r', 'd', 'b'}
7 >>> a | b # letters in a or b or both
8 {'a', 'c', 'r', 'd', 'b', 'm', 'z', 'l'}
9 >>> a & b # letters in both a and b
10 {'a', 'c'}
11 >>> a ^ b # letters in a or b but not both
12 {'r', 'd', 'b', 'm', 'z', 'l'}
```

• Từ điển là một bộ không có thứ tự, có thể thay đổi được và có chỉ số. Trong Python, một từ điển được viết trong cặp ngoặc nhọn và mỗi phần tử trong từ điển là một cặp gồm khóa và giá trị.

 Truy cập các phần tử trong từ điển
 Giá trị của mỗi phần tử trong từ điển có thể được lấy thông qua khóa của nó.

Giá trị của một khóa cũng có thể được lấy thông qua phương thức get()

```
x = thisdict.get("model")
```

Thay đổi giá trị

```
thisdict["year"] = 2019
```

Xâu ký tự - String

Duyệt các phần tử trong từ điển
 Dùng vòng lặp để duyệt qua các khóa trong từ điển

```
for x in thisdict:
    print(x)
# các khóa của từ điển ("brand", "model", "year")
```

Lấy từng giá trị trong từ điến

```
for x in thisdict:
print(thisdict[x])
```

 Sử dụng phương thức values() để lấy ra danh sách các giá trị có trong từ điển

```
for x in thisdict.values():
    print(x)
```

 Sử dụng phương thức items() lấy các phần tử trong từ điển là các cặp khóa – giá trị

 Sử dụng phương thức items() lấy các phần tử trong từ điển là các cặp khóa – giá trị

```
for x, y in thisdict.items():
    print(x, y)
```

Kiểm tra xem khóa có trong từ điển hay không

```
if "model" in thisdict:
    print("Yes, 'model' is one of the keys in the
    thisdict dictionary")

# Kích thước từ điển
print(len(thisdict))

# Thêm phần tử vào từ điển
thisdict["color"] = "red"
print(thisdict)
```

Xóa phần tử khỏi từ điển

```
# Xóa phần tử theo khóa bằng phương thức pop()
thisdict.pop("model")
print(thisdict)
# Xóa phần tử ngẫu nhiên bằng phương thức popitem()
thisdict.popitem()
# Xóa phần tử theo khóa bằng phương thức del
del thisdict["model"]
# Xóa cả từ điển bằng phương thức del
del thisdict
# Xóa trống từ điển bằng phương thức clear()
thisdict.clear()
```

Tạo bản sao của từ điển

```
# Tạo bản sao của từ điển bằng phương thức copy()
        thisdict = {
2
         "brand": "Ford",
3
         "model": "Mustang",
4
5
         "year": 1964
6
7
        mydict = thisdict.copy()
        print(mydict)
8
  # Một cách khác để tạo bản sao của từ điển là
  # sử dụng phương thức dict()
        mydict = dict(thisdict)
11
```

Từ điển lồng nhau

```
myfamily = {
    "child1" : {
       "name" : "Emil",
3
      "year" : 2004
4
    },
5
    "child2" : {
6
       "name" : "Tobias",
       "year" : 2007
8
    },
9
     "child3" : {
       "name" : "Linus",
11
       "year" : 2011
12
13
14 }
```

Xâu ký tự - String

Từ điển lồng nhau

```
child1 = {
     "name" : "Emil",
    "year" : 2004
3
4
  child2 = {
     "name" : "Tobias",
6
    "year" : 2007
7
8
  child3 = {
     "name" : "Linus",
10
    "year" : 2011
11
12
13
  myfamily = {
     "child1" : child1,
     "child2" : child2,
     "child3" : child3
17
18 }
19
```

Hàm dựng của từ điển

```
thisdict = dict(brand="Ford", model="Mustang", year=1964)
# note that keywords are not string literals
# note the use of equals rather than colon for the
    assignment
print(thisdict)
>>> {'brand': 'Ford', 'model': 'Mustang', 'year': 1964}
```

Xâu ký tự - String

```
1 >>> tel = {'jack': 4098, 'sape': 4139}
2 >>> tel['guido'] = 4127
3 >>> tel
4 {'jack': 4098, 'sape': 4139, 'guido': 4127}
5 >>> tel['jack']
6 4098
7 >>> del tel['sape']
8 >>> tel['irv'] = 4127
9 >>> tel
10 {'jack': 4098, 'guido': 4127, 'irv': 4127}
11
12 >>> list(tel)
13 ['jack', 'guido', 'irv']
14 >>> sorted(tel)
15 ['guido', 'irv', 'jack']
16 >>> 'guido' in tel
17 True
18 >>> 'jack' not in tel
19 False
```

```
1 >>> dict([('sape', 4139), ('guido', 4127), ('jack', 4098)])
2 {'sape': 4139, 'guido': 4127, 'jack': 4098}
3
|4| >>> \{x: x**2 \text{ for } x \text{ in } (2, 4, 6)\}
5 {2: 4, 4: 16, 6: 36}
6
7 >>> dict(sape=4139, guido=4127, jack=4098)
8 {'sape': 4139, 'guido': 4127, 'jack': 4098}
9
10 >>> knights = {'gallahad': 'the pure', 'robin': 'the brave'}
11 >>> for k, v in knights.items():
  ... print(k, v)
13
  . . .
14 gallahad the pure
15 robin the brave
```

```
# ví dụ về enumerate (liệt kê)

>>> for i, v in enumerate(['tic', 'tac', 'toe']):
... print(i, v)

...

0 tic
1 tac
2 toe
```

```
# Ham zip() de duyet tren nhieu chuoi dong thoi

>>> questions = ['name', 'quest', 'favorite color']

>>> answers = ['lancelot', 'the holy grail', 'blue']

>>> for q, a in zip(questions, answers):

... print('What is your {0}? It is {1}.'.format(q, a))

...

What is your name? It is lancelot.

What is your quest? It is the holy grail.

What is your favorite color? It is blue.
```

```
# Lăp theo thứ tự đảo ngược
>>> for i in reversed(range(1, 10, 2)):
... print(i)
...
5 9
6 7
7 5
8 3
9 1
```

```
# Tạo một danh sách mới lưu kết quả (đơn giản và an toàn hơn)

>>> import math

>>> raw_data = [56.2, ('NaN'), 51.7, 55.3, 52.5, ('NaN'), 47.8]

>>> filtered_data = []

>>> for value in raw_data:

... if not math.isnan(value):

... filtered_data.append(value)

...

>>> filtered_data

[56.2, 51.7, 55.3, 52.5, 47.8]
```

```
# Tạo một danh sách mới lưu kết quả (đơn giản và an toàn hơn)

>>> import math

>>> raw_data = [56.2, ('NaN'), 51.7, 55.3, 52.5, ('NaN'), 47.8]

>>> filtered_data = []

>>> for value in raw_data:

... if not math.isnan(value):

... filtered_data.append(value)

...

>>> filtered_data

[56.2, 51.7, 55.3, 52.5, 47.8]
```

```
# So sánh theo bộ
(1, 2, 3) < (1, 2, 4)
[1, 2, 3] < [1, 2, 4]

'ABC' < 'C' < 'Pascal' < 'Python'
(1, 2, 3, 4) < (1, 2, 4)
(1, 2) < (1, 2, -1)
(1, 2, 3) == (1.0, 2.0, 3.0)
(1, 2, ('aa', 'ab')) < (1, 2, ('abc', 'a'), 4)
```