

vncoder.vn

Bài 13: Hàm String - Numpy trong Python

4-5 phút

Các hàm sau đây được sử dụng để thực hiện các hoạt động chuỗi được vector hóa trong các mảng kiểu dtype numpy.string_ hoặc numpy.unicode_. Chúng dựa trên các hàm chuỗi tiêu chuẩn trong thư viện tích hợp sẵn của Python.

1. numpy.char.add() :

Hàm này thực hiện nối chuỗi phần tử

Ví dụ :

```
import numpy as np
print 'Concatenate two strings:'
print np.char.add(['hello'], [' xyz'])
print '\n'

print 'Concatenation example:'
print np.char.add(['hello', 'hi'], [' abc', ' xyz'])
```

Kết quả :

```
Concatenate two strings:
['hello xyz']
```

```
Concatenation example:
```

```
['hello abc' 'hi xyz']
```

2. numpy.char.multiply()

Hàm này thực hiện nhiều phép nối.

```
import numpy as np
print np.char.multiply('Hello ', 3)
```

Kết quả :

3. numpy.char.center()

Hàm này trả về một mảng có độ rộng cần thiết để chuỗi đầu vào được căn giữa và đệm ở bên trái, bên phải với **fillchar**.

```
import numpy as np
# np.char.center(arr, width, fillchar)
print np.char.center('hello', 20, fillchar =
'*')
```

Kết quả :

4. numpy.char.capitalize()

Hàm này trả về bản sao của chuỗi với chữ cái đầu tiên được viết hoa.

```
import numpy as np
print np.char.capitalize('hello world')
```

Kết quả :

5. numpy.char.title()

Hàm này trả về phiên bản viết hoa tiêu đề của chuỗi đầu vào với chữ cái đầu tiên của mỗi từ được viết hoa.

```
import numpy as np
```

```
print np.char.title('hello how are you?')
```

Kết quả :

6. numpy.char.lower()

Hàm này trả về một mảng có các phần tử được chuyển đổi thành chữ thường. Được gọi là str.lower cho mỗi phần tử.

```
import numpy as np
print np.char.lower(['HELLO', 'WORLD'])
print np.char.lower('HELLO')
```

Kết quả :

7. numpy.char.upper()

Hàm này gọi hàm str.upper trên mỗi phần tử trong một mảng để trả về các phần tử mảng chữ hoa.

```
import numpy as np
print np.char.upper('hello')
print np.char.upper(['hello', 'world'])
```

Kết quả :

8. numpy.char.split()

Hàm này trả về danh sách các từ trong chuỗi đầu vào. Theo mặc định, một khoảng trắng được sử dụng làm dấu phân cách. Nếu không, ký tự phân tách được chỉ định được sử dụng để làm tràn chuỗi.

```
import numpy as np
print np.char.split('hello how are you?')
print np.char.split
('TutorialsPoint,Hyderabad,Telangana', sep =
',')
```

Kết quả :

```
['hello', 'how', 'are', 'you?']  
['TutorialsPoint', 'Hyderabad', 'Telangana']
```

9. numpy.char.splitlines() :

Hàm này trả về danh sách các phần tử trong mảng, ngắt ở ranh giới dòng.

```
import numpy as np  
print np.char.splitlines('hello\nhow are you?')  
print np.char.splitlines('hello\rhow are you?')
```

Kết quả :

```
['hello', 'how are you?']  
['hello', 'how are you?']
```

'\n', '\r', '\r\n' có thể được sử dụng làm đường ranh giới.

10. numpy.char.strip()

Hàm này trả về một bản sao của mảng với các phần tử bị tước bỏ các ký tự đã chỉ định ở đầu và / hoặc ở cuối trong đó

```
import numpy as np  
print np.char.strip('ashok arora', 'a')  
print  
np.char.strip(['arora', 'admin', 'java'], 'a')
```

Kết quả :

```
shok aror  
['ror' 'dmin' 'jav']
```

11. numpy.char.join()

Phương thức này trả về một chuỗi trong đó các ký tự riêng lẻ

được nối với nhau bằng ký tự phân cách được chỉ định.

```
import numpy as np
print np.char.join(':', 'dmy')
print np.char.join([':', '-'], ['dmy', 'ymd'])
```

Kết quả :

12. numpy.char.replace()

Hàm này trả về một bản sao mới của chuỗi đầu vào, trong đó tất cả các lần xuất hiện của chuỗi ký tự được thay thế bằng một chuỗi đã cho khác.

```
import numpy as np
print np.char.replace ('He is a good boy',
                        'is', 'was')
```

Kết quả :

13. numpy.char.decode()

Hàm này gọi numpy.char.decode () giải mã chuỗi đã cho bằng cách sử dụng codec được chỉ định.

```
import numpy as np

a = np.char.encode('hello', 'cp500')
print a
print np.char.decode(a, 'cp500')
```

Kết quả :

14. numpy.char.encode()

Hàm này gọi hàm str.encode cho mỗi phần tử trong mảng. Mã hóa mặc định là utf_8, codec có sẵn trong thư viện Python chuẩn có thể được sử dụng.

```
import numpy as np
a = np.char.encode('hello', 'cp500')
print a
```

Kết quả :