

正規表現

正規表現	意味
?	0か1の文字
*	0以上の繰り返し
+	1以上の繰り返し
{}	複数回繰り返す
()	括弧
	並列
^	先頭
[]	範囲
a-z	小文字
A-Z	大文字
0-9	数字

正規表現	意味
\d	数字
\w	アルファベット
\s	空白
\D	数字以外
\W	アルファベット以外
\S	空白以外
.	任意の文字
^	先頭
\$	先端

match

正規表現と対象文字列を比較する
None

```
import re

re.match(pattern,string)
## pattern 正規表現
## string 対象文字列
```

```
a = re.match(r'test','testasdttest')
print(a)                                # None
print(a.group())                         # None
print(a.span())                          # None
print(re.match(r'test','atestasdttest'))    # None
```

search

検索する

```
import re

## search()match
a = re.match(r'test','atestasdttest')
b = re.search(r'test','atestasdttest')
print(a)                      # None
print(b)                      # <_sre.SRE_Match object at 0x00000000>
```

findall

複数の検索結果を取得する

- `re.s`属性

```
import re

result = re.findall(r'test','123test123test')
print(result)                  # ['test','test']

## テスト
a = """aaatestaa
aaaa123"""
print(re.findall(r'test',a))      # None
print(re.findall(r'test',a,re.s))    # ['aaatestaa']
```

sub

置換する

```
import re

## sub(置換する)
result = re.sub('php','python','phpwwwwww--php')
print(result)                  # "pythonwww--python"
```

split

分割する

```
import re

s = "itcast,java:php-php3;html"
print(re.split(r',',s))          # [,]
['itcast','java:php-php3;html']
print(re.split(r',|-|;',s))    # [,,:;-;]
['itcast','java','php','php3','html']
print(re.split(r',|-|%',s))   # []
['itcast','java','php','php3;html']
```

問題

- 正規表現の複数マッチ
- 複数マッチする場合、何が返る？

```
import re

text = 'abc123def456'
## []
match = re.search(r'\d+.*',text)
print(match.group())      # abc123def456
## []
match = re.search(r'\d+.*?',text)
print(match.group())      # abc123

## []
text = "<div>content</div><p>paragraph</p>"
match = re.search(r'<.*>', text)
print(match.group())      # <div>content</div><p>paragraph</p>
## []
match = re.search(r'<.*?>', text)
print(match.group())      # <div>
```