

Python URL エンコード

[Python]

- <http://d.hatena.ne.jp/niiyan/20090509/1241884365>
- <http://hoge.jp/blog/memo/2007/11/pythonurlliburlencode.html>

urllib

■ urllib.quote

構文

```
quote(s, safe='/')
```

- safe にエンコードしない文字を指定
- safe で指定しなければ、すべてエンコードする
- RFC 2396 Uniform Resource Identifiers (URI): Generic Syntax での予約、非予約文字

予約文字

```
reserved = gen-delims / sub-delims
gen-delims = ":" / "/" / "?" / "#" / "[" / "]" / "@"
sub-delims = "!" / "$" / "&" / "'" / "(" / ")"
              / "*" / "+" / "," / ";" / "="
```

非予約文字

```
unreserved = ALPHA / DIGIT / "-" / "." / "_" / "~"
```

区切り文字

```
delims = "<" / ">" / "#" / "%" / "<"
```

安全でない

```
unwise = "{" / "}" / "|" / "\" / "^" / "[" / "]" / "`"
```

例 1

```
>>> import urllib
>>> urllib.quote('abc def')
'abc%20def'
```

例 2

```
>>> urllib.quote('http://test.com?p=abc def')
'http%3A//test.com%3Fp%3Dabc%20def'
```

例 3

```
>>> urllib.quote('http://test.com?p=abc def', safe=';/?:@&=+$,')
'http://test.com?p=abc%20def'
```

■ urllib.urlencode

- ・ タプル、またはディクショナリを、URL クエリー文字列として生成する

```
>>> import urllib
>>> param = {'q': 'test', 'id': 'piroto'}
>>> urllib.urlencode(param)
'q=test&id=piroto'
```

HTML 特殊文字のエスケープ、アンエスケープ

- ・ xml.sax.saxutils の escape, unescape を利用すると便利

```
>>> from xml.sax.saxutils import *
>>> escape("<&")
'&lt;&amp;&gt;'
>>> unescape("&lt;&amp;&gt;")
'<&gt;'
>>> unescape("&#39;", {"&#39;":""})
''
```