

Regular Expressions in Python

L555

Dept. of Linguistics, Indiana University
Fall 2010

Module

In order to use regular expressions, we need to load the module.

```
import re
```

Regular Expression Symbols

.	wildcard
\	escapes special characters
[...]	character set
[^...]	complement of character set
	or
*	Kleene star
+	Kleene plus
{ m,n }	repeat between m and n times
^	beginning of a string
\$	end of a string

<code>compile(<pattern>)</code>	compiles a regex pattern into a pattern object – for reuse
<code>search(<pattern>, <string>)</code>	searches for regex pattern in string
<code>match(<pattern>, <string>)</code>	checks at beginning of string
<code>split(<pattern>, <string>)</code>	splits the string based on pattern, returns a list
<code>findall(<pattern>, <string>)</code>	returns a list of all occurrences
<code>sub(<pat>, <rep>, <string>)</code>	replaces pat by rep in string

Example

```
import re
```

```
mysent = raw_input('Gimme a sentence!')  
if (not re.search('[\s!\.;\?]', mysent)):  
    print 'this is not a sentence'
```

Example

```
import re
```

```
mysent = raw_input('Gimme a sentence!')
```

```
newstr = re.sub('[A-Z]', 'XX', mysent)
```

```
print newstr
```

Pattern Objects

Module

The functions `compile`, `search`, and `match` return a pattern object. The objects contain information about the pattern itself and for the matching functions also information about the matched segments in the string.

```
import re
```

```
phoneNums = re.compile('^[()?[0-9]{3,3}[- ]...  
....[0-9]{3,3}[- ]?[0-9]{4,4}$')
```

```
myphone = raw_input('Give me a phone number: ')  
if phoneNums.search(myphone):  
    print 'format correct'  
else:  
    print 'format incorrect'
```

Example

```
import re
```

```
mysent = ' a _rose _is _a _rose _is _a _rose '  
allstr = re.search('( _._ )', mysent)  
print allstr.group(1)
```