

example_re2

February 2, 2023

1 Example with re2

wraplib wraps the library *re2* using the wrapper *pyre2*.

```
[1]: from jupyterhelper import add_notebook_menu
      add_notebook_menu()
```

```
[1]: <IPython.core.display.HTML object>
```

```
[2]: from wraplib import re2
```

1.1 Example with HTML

```
[3]: import re
      s = "<h1>mot</h1>"
      print(re.compile("<.*>").match(s).groups())
```

```
('<h1>mot</h1>',)
```

```
[4]: s = "<h1>mot</h1>"
      print(re2.compile("<.*>").match(s).groups())
```

```
('<h1>mot</h1>',)
```

1.2 Group, Span

```
[5]: s = """date 0 : 14/9/2000
      date 1 : 20/04/1971    date 2 : 14/09/1913    date 3 : 2/3/1978
      date 4 : 1/7/1986     date 5 : 7/3/47     date 6 : 15/10/1914
      date 7 : 08/03/1941   date 8 : 8/1/1980   date 9 : 30/6/1976"""

      expression = re2.compile(
          "([0-3]?[0-9]/[0-1]?[0-9]/([0-2][0-9])?[0-9][0-9])[^\d]")
      expression.search(s).group(1, 2)
```

```
[5]: ('14/9/2000', '20')
```

```
[6]: c = expression.search(s).span(1)
      s[c[0]:c[1]]
```

```
[6]: '14/9/2000'
```

1.3 Names

```
[7]: date = "05/22/2010"
exp = "(?P<jj>[0-9]{1,2})/(?P<mm>[0-9]{1,2})/(?P<aa>((19)|(20))[0-9]{2})"
com = re2.compile(exp)
print(com.search(date).groupdict())
```

```
{'aa': '2010', 'jj': '05', 'mm': '22'}
```

1.4 findall

findall is not natively implemented in *re2*. It was added.

```
[8]: s = """date 0 : 14/9/2000
date 1 : 20/04/1971      date 2 : 14/09/1913      date 3 : 2/3/1978
date 4 : 1/7/1986       date 5 : 7/3/47        date 6 : 15/10/1914
date 7 : 08/03/1941     date 8 : 8/1/1980     date 9 : 30/6/1976"""

expression = re2.compile(
    "([0-3]?[0-9]/[0-1]?[0-9]/([0-2][0-9])?[0-9][0-9])[^\d]")

re2.findall(expression, s)
```

```
[8]: [('14/9/2000', '20'),
      ('20/04/1971', '19'),
      ('14/09/1913', '19'),
      ('2/3/1978', '19'),
      ('1/7/1986', '19'),
      ('7/3/47', None),
      ('15/10/1914', '19'),
      ('08/03/1941', '19'),
      ('8/1/1980', '19')]
```

1.5 benchmark

```
[9]: s = """date 0 : 14/9/2000
date 1 : 20/04/1971      date 2 : 14/09/1913      date 3 : 2/3/1978
date 4 : 1/7/1986       date 5 : 7/3/47        date 6 : 15/10/1914
date 7 : 08/03/1941     date 8 : 8/1/1980     date 9 : 30/6/1976"""

expression = re.compile(
    "([0-3]?[0-9]/[0-1]?[0-9]/([0-2][0-9])?[0-9][0-9])[^\d]")

%timeit expression.findall(s)
```

10.5 μ s \pm 296 ns per loop (mean \pm std. dev. of 7 runs, 100000 loops each)

```
[10]: %timeit re2.findall(expression, s)
```

18.4 μs \pm 1.51 μs per loop (mean \pm std. dev. of 7 runs, 10000 loops each)

That's expected as method *findall* is implemented in python and not C.

[11]: