プログラミングA2 第8回

練習問題解答例

練習問題:read_html.py

HTTPResponseオブジェクトはIterableであるため, for-in文で1行ずつhtmlソースを読み込むことができる.

for-in文で1行ずつ読み込み,改行文字を削除後に空である行を除いてprintするコードを実装せよ.

また, コンテキストマネージャ(with構文)を用いること.

```
read_html.py
url = "https://www.teu.ac.jp/"
with urlopen(url) as res:
    for row in res:
        row = row.decode().rstrip()
        if row == "":
            continue
        print(row)
```

- ← 改行文字削除後に
- ← 空行だったら

練習問題:poke_regexp2.py

「ュウ」または「ュー」で終わる名前を検索し、マッチした場合のみ名前とMatchオブジェクトを表示せよ

```
titles = read_names("lec08/data/poke_names.txt")
pattern = re.compile("(ュウ|ュー)$")
for i, title in enumerate(titles, 1):
    match = pattern.search(title)
    if match:
        print(title, match)
```

ピカチュウ <re.Match object; span=(3, 5), match='ュウ'> ライチュウ <re.Match object; span=(3, 5), match='ュウ'> ミニリュウ <re.Match object; span=(3, 5), match='ュウ'> ハクリュー <re.Match object; span=(3, 5), match='ュー'> カイリュー <re.Match object; span=(3, 5), match='ュー'> ミュウ <re.Match object; span=(1, 3), match='ュー'> ピチュー <re.Match object; span=(2, 4), match='ュー'> デンリュウ <re.Match object; span=(3, 5), match='ュウ'>

練習問題:poke_regexp2.py

以下の正規表現を記述し、正しくマッチできるか確認せよ.

●「ン」を2回含む名前

```
pattern = re.compile("\vee.*\vee")
```

```
サンドパン <re.Match object; span=(1, 5), match='ンドパン'>
コンパン <re.Match object; span=(1, 4), match='ンパン'>
ランターン <re.Match object; span=(1, 5), match='ンターン'>
ヤンヤンマ <re.Match object; span=(1, 4), match='ンヤン'>
アンノーン <re.Match object; span=(1, 5), match='ンノーン'>
マンタイン <re.Match object; span=(1, 5), match='ンタイン'>
ドンファン <re.Match object; span=(1, 5), match='ンファン'>
```

●「ラリルレロ」のいずれかを2個以上含む名前

pattern = re.compile("[ラリルレロ].*[ラリルレロ]")

```
トランセル <re.Match object; span=(1, 5), match='ランセル'>
オニドリル <re.Match object; span=(3, 5), match='リル'>
ラフレシア <re.Match object; span=(0, 3), match='ラフレ'>
オコリザル <re.Match object; span=(2, 5), match='リザル'>
レアコイル <re.Match object; span=(0, 5), match='リアコイル'>
ビリリダマ <re.Match object; span=(1, 3), match='リリ'>
カラカラ <re.Match object; span=(1, 4), match='ラカラ'>
:
デルビル <re.Match object; span=(1, 4), match='ルビル'>
```

↑ 最後に「.*」を付けてみると matchがどう変わる?

練習問題:poke_regexp2.py

以下の正規表現を記述し、正しくマッチできるか確認せよ.

●3文字の名前

```
pattern = re.compile("^...$") または "^.{3}$"

ポッポ <re.Match object; span=(0, 3), match='ポッポ'>
ラッタ <re.Match object; span=(0, 3), match='ラッタ'>
アーボ <re.Match object; span=(0, 3), match='アーボ'>
:
ブビィ <re.Match object; span=(0, 3), match='ブビィ'>
ルギア <re.Match object; span=(0, 3), match='ルギア'>
```

●小さな「ッ」と任意の1文字で終わる名前

```
pattern = re.compile("ッ.$")
```

```
ポッポ <re.Match object; span=(1, 3), match='ッポ'>
ピジョット <re.Match object; span=(3, 5), match='ット'>
コラッタ <re.Match object; span=(2, 4), match='ッタ'>
ラッタ <re.Match object; span=(1, 3), match='ッタ'>
アーボック <re.Match object; span=(3, 5), match='ック'>
こ
エレキッド <re.Match object; span=(3, 5), match='ッド'>
```

練習問題:html_regexp1.py

前ページのURLのうち、大学学外へのもののみprintするように変更せよ.

```
html_regexp1.py
html = read_html("lec08/data/index.html")
pattern = re.compile("<a href=\frac{1}{2}"(http.*?)\frac{1}{2}".*?>(.*?)</a>")
match_lst = pattern.findall(html)
for match in match_lst:
    print(match)
```

```
:
('http://twitter.com/tut_tweet', '<span>twitter</span>')
('http://www.youtube.com/user/TokyoKokaDaigaku',
'<span>YouTube</span>')
('http://line.naver.jp/ti/p/%40koukadai', '<span>LINE</span>')
('https://page.line.me/koukadai', '<span>LINE</span>')
('https://www.instagram.com/tut_koukaton.official/',
'<span>instagram</span>')
```

練習問題:html_regexp2.py

<!--と --> で囲まれたコメントアウト部分を削除するコードを実装せよ.

ただし、コメントアウトは複数行にまたがることが多いことに注意せよ.

```
html_regexp2.py

html = read_html("lec08/data/index.html").replace("¥n", "")

pattern = re.compile("<!¥-¥-.*?¥-¥->")

match_lst = pattern.findall(html) ← 正規表現が ← 正しいか ← 確認

html = pattern.sub("", html)
```