

الدرس دا هيكون اخر دروس ال **Regular Expressions** وهنتكلم فيه عن ال groups وال .flags

موجود في المثال عندي pattern بيعمل match ل URL وال pattern معمول بطريقة ال .groups

انا عامل في ال pattern شوية groups كل واحد بيعمل match لحاجه في ال URL، يعني:

- ال group الاول مخليه يعمل match لجزء ال protocol اللي هو بيكون http او https او ftp او غيرهم من ال protocols ودا ممكن يكون موجود في ال URL وممكن ميكونش موجود علشان كدا حظيت علامة ؟ .
- ال group الثاني بيعمل match لجزء ال Sub Domain ودي اللي بتكون www ودي برضو احيانا بتكون موجوده في ال URL واحيانا مش بتكون موجوده علشان كدا برضو مستخدمين علامة ؟ .
- ال group الثالث ودا بيعمل match لل Domain Name بتاعي اللي هو بيكون اسم الموقع زي google او Facebook وغيرهم، ودا لازم يكون موجود في ال URL اساسي علشان كدا محطيناش بعد علامة ؟ .
- ال group الرابع ودا بيعمل match لل Top level domain ودا اللي بيكون .com او .org او .net وغيرهم، ودا لازم يكون موجود برضو علشان كدا محطيناش علامة ؟ .

بعد ال Top level domain المفروض بيكون كدا ال URL انتهى، لكن احيانا لما بيكون في request بيتم فيبتكتب في ال URL رقم ال port وال query string فعملتلهم groups وحظيت بعدهم علامة ؟ علشان احتمال يكونوا موجودين واحتمال لا.

- ال group الخامس بيعمل match لرقم ال port ودا بيكون ارقام فبكتبه كدا (\d+) وعلشان ممكن يكون موجود وممكن يكون لا فبخط بعده علامة ؟ .
- ال group السادس بيعمل match ل ال query string ودا بيكون كله عبارته عن حروف وارقام بقا فبقله حددي كل حاجه سواء حرف او رقم عن طريق (+). ودا كمان بيكون في احتمال لوجوده او لا فبستخدم علامة ؟ .

عندنا بقا 2 فانكشنز بستخدمهم علشان اقدر اتعامل مع ال groups وهم **group()** و **.groups()**

group()

بستخدمها علشان اعرض ال groups اللي عملتلها match.

ال argument بخط فيها رقم ال group اللي عايز اعرضه ولو سيبتها فاضيه بترجعلي ال string كله.

يعني في المثال بتاع ال URL لو عايز اعرض ال protocol اللي هو اول group بكتب في ال argument 1 كدا group(1) ولو عايز الثاني بتاع ال sub domain بكتب 2 وهكذا.

groups()

بتعرضلي كل ال groups اللي حصلها match في ال string وبتحطهملي في tuple واقدر اني اعمل loop ع ال list دي واخرج كل ال groups اللي حصلها match واعمل عليهم فانكشنز او اي حاجه.

نيجي بقا لجزئية ال flags ودي اللي بتدي خصائص لعملية ال matching بتاعتي لما بستخدم فانكشن `re.search()`. بكل بساطه بقدر اني ازود فيها argument كمان واكتب فيه ال flag بتاعي عن طريق اني بكتب `re.` وبعدها ال flag اللي عايز احطه.

يعني على سبيل المثال لو عملت **`re.I`** دا flag بيعملي ignore case في عملية ال matching يعني بيهمل حالة الحرف اذا كان capital او small.

عندي كمان **`re.M`** ودي اختصار ل multi line وبتعملي match لكل سطور ال string لو هو مكتوب على اكثر من سطر. في غيرهم flags ممكن تبقا تدور عليهم.

```
# -----
# -- Regular Expressions => Group Trainings And Flags --
# -----

import re

my_web = "https://www.elzero.org:8080/category.php?article=105?name=how-to-do"

search = re.search(r"(https?)://(www)?\.(?(\w+)\.(\w+):?(\d+)?/?(.+)", my_web)

print(search.group())
print(search.groups())

for group in search.groups():

    print(group)

print(f"Protocol: {search.group(1)}")
print(f"Sub Domain: {search.group(2)}")
print(f"Domain Name: {search.group(3)}")
print(f"Top Level Domain: {search.group(4)}")
print(f"Port: {search.group(5)}")
print(f"Query String: {search.group(6)}")

my_string = "MUHAMmAd AHMed"
my_search = re.search(r"[A-Z]{8}", my_string) # without flag: None
print(my_search)

my_string = "MUHAMmAd AHMed"
my_search = re.search(r"[A-Z]{8}", my_string, re.I) # with flag: match found
print(my_search)
```

CODE

```
https://www.elzero.org:8080/category.php?article=105?name=how-to-do
('https', 'www', 'elzero', 'org', '8080', 'category.php?article=105?name=how-to-do')
https
www
elzero
org
8080
category.php?article=105?name=how-to-do
Protocol: https
Sub Domain: www
Domain Name: elzero
Top Level Domain: org
Port: 8080
Query String: category.php?article=105?name=how-to-do
None
<re.Match object; span=(0, 8), match='MUHAMMAd'>
```

OUTPUT