

Automatyzacja nudnych zadań w Pythonie

Rozdział 7

1. Która funkcja tworzy obiekty Regex?
re.compile()

2. Dlaczego niezmodyfikowane ciągi tekstowe są często używane podczas tworzenia obiektów Regex?

Niezmodyfikowane ciągi tekstowe są używane, aby nie trzeba było stosować znaków sterujących dla ukośników.

3. Co jest wartością zwrótną metody search()?
obiekt Match lub NoneType

4. Jak z obiektu Match można otrzymać rzeczywisty ciąg tekstowy, który dopasował wzorzec?
za pomocą metody group() lub group(0)

5. Co zostanie dopasowane przez grupy 0, 1 i 2 w podanym wyrażeniu regularnym: r'(\d\d\d)-(\d\d\d-\d\d\d\d)'?

dla grupy 0 - '(\d\d\d)-(\d\d\d-\d\d\d\d)'

dla grupy 1 - '(\d\d\d)'

dla grupy 2 - '(\d\d\d-\d\d\d\d)'

6. Nawias i gwiazdka mają znaczenie specjalne w składni wyrażeń regularnych. Jak mógłbyś utworzyć wyrażenie regularne, które miałoby dopasować dosłowny nawias lub kropkę?

nawias - \(\)

kropka - \.

7. Metoda findall() zwraca listę ciągów tekstowych lub listę krotek ciągów tekstowych. Dlaczego tak się dzieje?

Jeżeli wyrażenie regularne nie ma grup, zwrócona będzie lista ciągów tekstowych. Jeśli natomiast wyrażenie regularne ma grupy, wówczas zwrócona będzie lista krotek ciągów tekstowych.

8. Jakie jest znaczenie znaku | w wyrażeniach regularnych?

Jest to dopasowanie jedno z wielu wyrażeń. (Dopasuj to lub to)

9. Jakie dwa znaczenia w wyrażeniach regularnych ma znak zapytania?

1. Dopasowanie zero lub jednej grupy

2. Dopasowanie niezachłanne.

10. Na czym polega różnica między znakami plus a gwiazdki w wyrażeniach regularnych?

znak + jest dopasowaniem jednego lub więcej wyrażeń (jeżeli nie znajdzie chociaż jednego dopasowania to zwróci obiekt NoneType)

znak * oznacza dopasowanie typu jeden lub więcej

znak * jest dopasowaniem zero lub więcej wyrażeń.

11. Na czym polega różnica między {3} i {3,5} w wyrażeniach regularnych?

przykładowo:

(\d){3} dopasuje trzy cyfry,

(\d){3,5} dopasuje 3, 4 lub 5 cyfr

12. Jakie znaczenie w wyrażeniach regularnych mają skróty klas znaków \d, \w i \s?

\d - dopasowuje tylko cyfry

\w - dopasowuje litery, cyfry lub znak podkreślenia

\s - spacje, tabulator lub znak nowego wiersza

13. Jakie znaczenie w wyrażeniach regularnych mają skróty klas znaków \D, \W i \S?

\D - dopasowuje wszystko oprócz cyfr

\W - dopasowuje wszystko oprócz liter, cyfr lub znaków podkreślenia

\S - dopasowuje wszystko prócz spacji, tabulatorów lub znaków nowego wiersza

14. Na czym polega różnica między .* a .*??

.* jest dopasowaniem zachłannym

.*? jest dopasowaniem niezachłannym

15. Jak przedstawia się składnia klasy znaków dopasowującej wszystkie cyfry i małe litery?

[a-z0-9]

16. Co można zrobić, aby wyrażenia regularne nie uwzględniały wielkości liter?

stosując re.I lub re.IGNORECASE

17. Co standardowo dopasowuje kropka? Co będzie dopasowane przez kropkę, gdy zmienna re.DOTALL zostanie przekazana jako drugi argument funkcji re.compile()?

Standardowo kropka dopasowuje dowolny znak z wyłączeniem znaku nowego wiersza.

Jeżeli zostanie przekazane re.DOTALL to kropka będzie dopasowywała wszystkie znaki razem z znakiem nowego wiersza.

18. Przyjmujemy założenie, że numRegex = re.compile(r'\d+'). Jaka będzie wartość zwrotna numRegex.sub('X', '12 drummers, 11 pipers, five rings, 3 hens')?

'X drummers, X pipers, five rings, X hens'

19. Jaki będzie skutek przekazania zmiennej re.VERBOSE jako drugiego argumentu funkcji re.compile()?

Zmienna re.VERBOSE pozwala na dodanie białych znaków i komentarzy do ciągu tekstowego przekazywanego funkcji re.compile().

20. Jakie utworzyłbyś wyrażenie regularne dopasowujące liczbę wraz z przecinkami po każdych trzech cyfrach? To wyrażenie musi dopasować poniższe liczby:

'42'

'1,234'

'6,368,745'

ale nie może dopasować następujących:

'12,34,567' (tylko dwie cyfry między przecinkami),

'1234' (brak przecinków).

```
re.compile(r'(\d(\,(\d){3})+)|(\d){1,3}')
```

21. Jakie utworzyłbyś wyrażenie regularne dopasowujące pełne imię i nazwisko osoby, której nazwisko to Watanabe? Możesz przyjąć założenie, że imię zawsze jest przed nazwiskiem, składa się z pojedynczego słowa i zaczyna wielką literą. Twoje wyrażenie regularne musi dopasować poniższe wartości:

'Haruto Watanabe'

'Alicja Watanabe'

'RoboCop Watanabe'

ale nie może dopasować następujących:

'haruto Watanabe' (pierwsza litera imienia jest mała),

'Mr. Watanabe' (słowo przed nazwiskiem ma znak inny niż litera),

'Watanabe' (brak imienia),

'Haruto watanabe' (pierwsza litera nazwiska jest mała).

```
re.compile(r'[A-Z]([a-zA-Z])* Watanabe')
```

22. Jakie utworzyłbyś wyrażenie regularne dopasowujące zdanie, w którym pierwsze słowo to Alicja, Bartek lub Celina, drugie słowo to je, głaszcze lub rzuca, a trzecie to słowo jabłko, kota lub piłkę, a zdanie kończy się kropką? To wyrażenie regularne nie powinno uwzględniać wielkości liter i musi dopasować poniższe zdania:

'Alicja je jabłko.'

'Bartek głaszcze kota.'

'Celina rzuca piłkę.'

'Alicja rzuca jabłko.'

'BARTEK GŁASZCZE KOTA.'
ale nie może dopasować następujących:
 'RoboCop je jabłko.'
 'ALICJA RZUCA PIŁKĘ.'
 'Celina je siódmego kota.'

```
re.compile(r'(Alicja|Bartek|Celina)+ (je|głaszcze|rzyca)+ (jabłko|kota|  
piłkę)+\.', re.I)
```