El?z? gyakorlat - Fel - Következ? gyakorlat

Tartalomjegyzék

- 1 Reguláris kifejezések
  - 1.1 Karakterosztályok
  - 1.2 Ismétl?dések
  - 1.3 Választás
  - 1.4 Csoportosítás
- 2 Feladatok
- 3 Házi feladat

Reguláris kifejezések

A reguláris kifejezéseket akkor használjuk, amikor összetettebb karaktersort keresünk egy szövegben, illetve ilyen karakterláncokat szeretnénk helyettesíteni. Az órán a https://regex101.com/#python oldalon fogjuk kipróbálni, hogy mi hogyan m?ködik.

- Speciális karakterek: nem ismeri fel ?ket, más jelentésük van, csak \ el?kével tudunk rájuk keresni, pl \$, \^ stb.

```
. ^ $ * + ? { } [ ] \ | ( )
```

Karakterosztályok

Egyel?re egyetlen karakterre illesztünk.

- **\d**: tetsz?leges számjegy, **\D**: tetsz?leges karakter, ami nem számjegy.
- **\w**: tetsz?leges alfanumerikus karakter, bet?, szám vagy alulvonás (_), **\W**: tetsz?leges, nem alfanumerikus karakter.
- **\s**: whitespace, azaz tab, sor vége, szóköz, **\S** tetsz?leges, nem whitespace karakter.
- Mi is megadhatunk karakterosztályt: **[xyz]**, vagy tilthatunk is, pl. **[^xyz]**. El?bbi az x,y,z karakterek bármelyikére illeszkedik, utóbbi bármilyen karakterre, ami nem x,y vagy z. S?t, köt?jellel elválasztva intervallunként hivatkozhatunk a karakterekre, pl. **[a-z]** illeszkedik az összes kisbet?re, de az **[A-Za-z0-9]** is helyes, ez minden kis- és nagybet?re és számjegyre illeszkedik.
- **^**: sor eleje, **$**, sor vége.
- A **.** minden karakterre illeszkedik.

Ismétl?dések

Jelölés	Ismétl?dések száma	Példa
*	0,1, vagy akármennyi	a **\d*** illeszkedik a '123' karakterekre, de az üres *-re is*
+	legalább 1	**\d+** Illeszkedik tetsz?leges darab számjegyre
?	0 vagy 1	**oda\-?vissza** illeszkedik az 'oda-vissza' és az 'odavissza' szavakra is
{m,n}	Legalább *m*, legfeljebb *n* darab, bármelyik szám elhagyható	**:D{4,10}** nem illeszkedik a ':DDDDDDDDDDDDD'-re

Választás

- Bármilyen magánhangzóra illeszkedik a **a|e|i|o|u** kifejezés. Próbáljuk ki a **GetValue|Get|Set|SetValue** kifejezést. A *SetValue* szövegnél mire illeszkedik?

Csoportosítás

Lehet?ségünk van csoportokat kijelölni a kifejezésen belül. Az alábbi példa olyan szövegre illeszkedik, amelyben kétszer ismétl?dik bármi.

```
(.*)\1
```

Tudunk keresni html tageket is

```
<([A-Z][A-Z0-9]*)\b[^>]*>.*?</\1>
```

Több csoportot is tudunk kezelni, a nyitó zárójelek sorrendje határozza meg a csoport sorszámát. Cseréljük le az email cím országkódját .hu-ra!

```
(\w+)@((\w+)\.)+(\w+)
```

Feladatok

yyyy.mm.dd formátumú dátumok

```
\d{4}\.\d{2}\.\d{2}
```

Magyar mobilszámok

```
06\-(3|2|7)0\-\d{7}
```

Link keresése html oldalon

```
<a.*\shref=\"http://wiki.math.bme.hu([^\"http://wiki.math.bme.hu>]+\"http://wiki.math.bme.hu).*>.*
```

Webcímek

```
 http://(www.)?([a-z0-9_]+\.)+(\w{2,3})
```

Keressük ki a hét képét a bmeme.hu oldalon!

Néggyel osztható 2-jegy? szám

```
[02468][048]|[13579][26]
```

Szök?év

```
([1-9][0-9])(0[48]|[2468][048]|[13579][26])|(([2468][048]|[13579][26])00)
```

Nagy bet?kkel írt római számok

Évezred: `M{0,4}`, évszázad: `CM|CD|D?C{0,3}`, évtized: `XC|XL|L?X{0,3}`, év: `IX|IV|V?I{0,3}`. Akkor mégis mi a hiba az alábbi megoldással?

```
M{0,4}(CM|CD|D?C{0,3})(XC|XL|L?X{0,3})(IX|IV|V?I{0,3})
```

Hogy illeszkedik az üres sztringre is! Megoldás

```
\b(?=[MDCLXVI])M{0,4}(CM|CD|D?C{0,3})(XC|XL|L?X{0,3})(IX|IV|V?I{0,3})\b
```

Pozitív egészek, esetleg a sokjegy? számok hármasával csoportosítva vannak és a csoportok szóközzel elválasztva

```
[1-9](\d{0,2}( \d{3})+|\d*)
```

Ugyanez szóhatárok közt:

```
\b[1-9](\d{0,2}( \d{3})+|\d*)\b
```

HTML-kódban hexadecimális színkód (3 vagy 6 hexa szám)

```
([0-9A-Fa-f]{3}){1,2}
```

Házi feladat

Itt található

El?z? gyakorlat - Fel - Következ? gyakorlat