

El?z? - Fel - Következ?

Tartalomjegyzék

- 1 Reguláris kifejezések
 - ◆ 1.1 Karakterosztályok
 - ◆ 1.2 Ismétlések
 - ◆ 1.3 Választás
 - ◆ 1.4 Csoportosítás
- 2 Feladatok
 - ◆ 2.1 Nehezebb feladatok

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 működik.

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

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

Karakterosztályok

Egyelőre egyetlen karakterre illesztünk.

- \d: tetszőleges számjegy, \D: tetszőleges karakter, ami nem számjegy.
- \w: tetszőleges alfanumerikus karakter, betű, szám vagy alulvonás (_), \W: tetszőleges, nem alfanumerikus karakter.
- \s: whitespace, azaz tab, sor vége, szóköz, \S tetszőleges, nem whitespace karakter.
- Mi is megadhatunk karakterosztályt: [xyz], vagy tilthatunk is, pl. [^xyz]. Előbbi az x,y,z karakterek bármelyikére illeszkedik, utóbbi bármilyen karakterre, ami nem x,y vagy z. Sőt, kötőjellel elválasztva intervallunként hivatkozhatunk a karakterekre, pl. [a-z] illeszkedik az összes kisbetűre, de az [A-Za-z0-9] is helyes, ez minden kis- és nagybetűre és számjegyre illeszkedik.
- ^: sor eleje, \$, sor vége.
- A . minden karakterre illeszkedik.

Ismétlések

Jelölés	Ismétlések száma	Példa
*	0,1, vagy akármennyi	a \d* illeszkedik a '123' karakterekre, de az üres <i>-re is</i>
+	legalább 1	\d+ Illeszkedik tetszőleges 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 ':DDDDDDDDDDDDDDDD'-re

Választás

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

Csoportosítás

Lehetőségünk van csoportokat kijelölni a kifejezésen belül. Az alábbi példa olyan szövegre illeszkedik, amelyben kétszer ismétlődik 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 országcódját .hu-ra!

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

Feladatok

- yyyy.mm.dd formátumú dátumok
- Magyar mobilszámok
- Négyvel osztható 2-jegyű szám
- Szökőév
- yyyy.mm.dd formátumú dátumok
 - ◆ de az elválasztó más is lehet, nem csak pont, viszont a két elválasztónak azonosnak kell lennie.

yyyy.mm.dd
yyyy-mm-dd
yyyy,mm,dd

- replace-el:
 - ◆ alakítsuk át a yyyy-mm-dd formátumú dátumokat mm-dd-yy alakúra!
 - ◆ cseréljük fel két oszlopot ebben: `[1]`

Nehezebb feladatok

Nagy betűkkel 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

Informatika1-2019/Gyakorlat3

`\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 sokjegyű számok hármásával 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}`

Előző - Fel - Következő