

REGEX pavyzdys su skirtingais lietuviškais telefono numeriais

Paveikslėlio apačioje matomi penki lietuviški numeriai aprašyti skirtingais formatais.

Pattern

```
(\\+370|8)[ -]?\\((?\\d{3})\\)?[ -]?\\d{5}
```

Input

```
+370 604 06645
+370 (604) 00654
+370-604-00654
+37060400654
860400654
```

Naudojantis *reguliariosiomis išraiškomis* (angl. REGEX) tekste galime atrasti šiuos skirtingus telefono numerius tekste. Panaudojimo pavyzdys: nustatyti, kiek kartų tekste buvo parodomas lietuviškas telefono numeris. Pavyzdys pateikiamas tolesniame paveikslėlyje:

Pattern

```
(\\+370|8)[ -]?\\((?\\d{3})\\)?[ -]?\\d{5}
```

Input

```
Jeif telefono numerio formatas yra toks +370 604 06645, tada būtų verta
susimastyti apie šiek tiek kitoki. +370 (604) 00654. Žinant tokių
telefono numerių formato populiarumą: +370-604-00654, neatmetama, kad
tokie numeriai taipogi neatsilieka:+37060400654. O 860400654 garbingasis
ir visiems įprastas yra laikomas populiariausiu
|
```

Naudotas įrankis: [REGEX Storm](#)

Paaiškinimai

(\+370|8) [-]? \ (? \d{3} \) ? [-]? \d{5}

Ši reguliarioji išraiška turi išskirtas kelias skirtingas grupes, pasirenkančias skirtingus simbolius.

Geltonai pažymėta dalis: (\+370|8)

() - Skliausteliais apgliaubta dalis pavirsta grupe. Grupė – tai konkrečiai simbolių grupei atrasti apibrėžiama išraiška. Grupės išraiška nežymi kitų grupių simbolių. O grupėje esanti išraiška nusako, kad šioje grupėje gali būti apibrėžti keli galimi variantai numerio pradžiai. Lietuviškas numeris prasideda dviejais variantais: +370 arba 8. Toliau einanti dalis daugeliu atvejų išlieka tokia pati: 6***** (* - bet koks skaičius)

\+: apibrėžia įprastą plius simbolį. (kitu atveju tai yra skaitoma kaip specialus simbolis – *1 arba daugiau*). Kitais atvejais + simbolis yra specialusis simbolis ir tai nėra skaitoma kaip būsimo teksto dalis. Atvirkštinis pakrypęs brūkšnelis (angl. *Backslash*) yra specialusis simbolis apverčiantis specialųjį simbolį einantį už jo į paprastą simbolį arba paprastą simbolį paverčia specialiuoju simboliu. (PVZ *\t reiškia tabuliacinės žymės simbolį*)

370 - įprasta simbolių seka, šiuo atveju viena iš galimų variantų grupėje. Sujungus praeitą ir šią išraišką. Jau galima atrasti lietuviškų numerių pradžias, prasidedančias +370.

| - Arba. Naudojama apibrėžti du arba daugiau leidžiamų variantų toje pačioje grupėje arba išraiškoje. | yra specialusis simbolis

8 – Šiuo atveju tai yra antroji galimybė vietoje +370;

Žaliai pažymėtos dalys: [-]?

[] – Laužtiniuose skliaustuose apibrėžiama dalis tampa vieno simbolio grupe. Juose aprašomi visi galimi simboliai einantys toje simbolio (Pabrėžiama **vienaskaita**) grupėje. Čia gali būti išvardinami simboliai tokiu principu: **[abcdefghiABCDEFGHI]** arba **[a-zA-Z]**. Tokiu atveju ši grupė ras visus simbolius nuo A mažosios iki I didžiosios ir mažosios. Verta paminėti, jog [ir] yra specialieji simboliai.

Tarpas ir brūkšnelis laužtiniuose skliaustuose: Tai reiškia kad yra du galimi simbolio variantai – tarpas ir brūkšnelis.

? – Klaustukas yra specialusis simbolis, sakantis jog toks **simbolis, simbolio grupė** arba **grupė** gali egzistuoti, tačiau gali ir neegzistuoti.

Pilkai pažymėta dalis: \ (? \d{3} \) ?

PASTABA: Šioje išraiškoje \ (\) dalis nėra grupė, kadangi (ir) simboliai yra pažymimi kaip paprastieji simboliai (*angliškas tikslus pavadinimas Escaped characters, toliau bus vadinama - eskeipinti*).

\(? Ir \)? – Šiuo atveju eskeipintas (simbolis su klaustuku reiškia kad tai yra įprastas skliaustelio simbolis kuris gali egzistuoti, tačiau gali ir neegzistuoti.

\d – bet koks skaičiaus simbolis

`\d{3}` - Trys bet kokio skaičiaus simboliai einantys iš eilės.

`\(?:\d{3}\)?` **Išvada:** pasirenka tris skaičius skliausteliuose arba ne skliausteliuose.

`\d{5}` **Išvada:** pasirenka penkis bet kokius skaičius iš eilės.

Tokiu principu su šia išraiška galime pasirinkti didžiąją daugumą lietuviškų numerių.

Užduotis, bei pasiūlymas išraiškos patobulinimui: *Patobulinkite išraišką kad aptiktų ir tokius telefono numerius: (+370) 6 00 00000*